

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:09:21 ; Search time 14 Seconds
(without alignments)
2882.455 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 775

Sequence: 1 MTIVDKASSSDPFSAVQNP.....AAESLEEDPAAASLPPFSEG 775

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 10

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	565	72.9	1198	1 UB42_HUMAN	Q9h9j4 homo sapien
2	14	1.8	1121	1 UB36_HUMAN	Q9p275 homo sapien

ALIGNMENTS

RESULT 1
UB42_HUMAN
ID UB42_HUMAN STANDARD; PRT; 1198 AA.
AC Q9H9J4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 42 (EC 3.1.2.15) (Ubiquitin
thiolesterase 42) (Ubiquitin-specific processing protease 42)
DE (Deubiquitinating enzyme 42) (Fragment).
GN USP42
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -1- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AK022759; BAB14232.1; -.
CC MEROPS: C19.048; -.
DR Genew; HGNC:20068; USP42.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; FALSE_NEG.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT_SITE 120 120 BY SIMILARITY.
FT ACT_SITE 362 362 BY SIMILARITY.
FT ACT_SITE 371 371 BY SIMILARITY.
FT NON_TER 1198 1198
SQ SEQUENCE 1198 AA; 130587 MW; CF0FDB5184AE9536 CRC64;

Query Match 72.9%; Score 565; DB 1; Length 1198;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 765; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPFSAVQNPQGSSEAVSPGMDAGSASWGAVSLNDVSNHTLSLGFVPGA 60
DB 1 MTIVDKASSSDPFSAVQNPQGSSEAVSPGMDAGSASWGAVSLNDVSNHTLSLGFVPGA 60

QY 61 VYSSSSVPDKSKPSQKQDQALGDGIAPQKVLFPSEKICLKWQTHRVGAGLQNLGNTC 120
DB 61 VYSSSSVPDKSKPSQKQDQALGDGIAPQKVLFPSEKICLKWQTHRVGAGLQNLGNTC 120

QY 121 FANAALQCLITYTPPLANYMLSHESKTCCHAEFCMCTMCAHITQALSNPQGVIKPMFVI 180
DB 121 FANAALQCLITYTPPLANYMLSHESKTCCHAEFCMCTMCAHITQALSNPQGVIKPMFVI 180

QY 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLS 240
DB 181 NEMRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLS 240

QY 241 RVKCLNCKGVSDTFDPLDITLLEIKAAQSVNKALQFVKPEQLDGENSYKSKCKMWPA 300
DB 241 RVKCLNCKGVSDTFDPLDITLLEIKAAQSVNKALQFVKPEQLDGENSYKSKCKMWPA 300

QY 301 SKRFTIHRSSNVLTSLKRFANFTGCKIAKQVPEYLDIRPYMOPNGEPIVYVLYAVL 360
DB 301 SKRFTIHRSSNVLTSLKRFANFTGCKIAKQVPEYLDIRPYMOPNGEPIVYVLYAVL 360

QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNDISVSTSDIRSVLSQQAAYVLFYIRSHDVKNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNDISVSTSDIRSVLSQQAAYVLFYIRSHDVKNGE 420

QY 421 LTHPTHSPQSSPRPVISORVVVTKQAAPGFIGPOLPSHMIKNPPLHNGTGLKDTPTSS 480
DB 421 LTHPTHSPQSSPRPVISORVVVTKQAAPGFIGPOLPSHMIKNPPLHNGTGLKDTPTSS 480

QY 481 MSSPNGNSVNRASPVNASASVQNVSNRNVSVIPEHPKKQKITIISHNKLPVRCQSQPN 540
DB 481 MSSPNGNSVNRASPVNASASVQNVSNRNVSVIPEHPKKQKITIISHNKLPVRCQSQPN 540

QY 541 LHSNSLENPTKVPFSSITNSAVQSTNSASTMSVSSKVTKEIPRSESCSQPVNMGSKLN 600
DB 541 LHSNSLENPTKVPFSSITNSAVQSTNSASTMSVSSKVTKEIPRSESCSQPVNMGSKLN 600

QY 601 SSVLPVPGAESSEDSDESKGLGKNGITVSSHSPQDAEDEATPEHLOEPMTLNGA 660
DB 601 SSVLPVPGAESSEDSDESKGLGKNGITVSSHSPQDAEDEATPEHLOEPMTLNGA 660

```
Db 601 SSVLPYGAESDSDSESKGLKENGITGVSSHSPGQDAEDEATPHELOPTTLNGA 660
661 NSADSDSPKENGGLAPDGASCOQOPALHSENPFKANGLPGLMPAPLLSLPDKKILETF 720
661 NSADSDSPKENGGLAPDGASCOQOPALHSENPFKANGLPGLMPAPLLSLPDKKILETF 720
QY 721 RLNKLKGTDEMSAPGAERGPEDRDABPQSPAAESLEEDPAAA 767
Db 721 RLNKLKGTDEMSAPGAERGPEDRDABPQSPAAESLEEDPAAA 767

RESULT 2
UB36_HUMAN STANDARD; PRT; 1121 AA.
AC Q9P275; Q8NDM8; Q9NVC8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 36 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 36) (Ubiquitin-specific processing protease 36)
DE (Deubiquitinating enzyme 36).
GN USP36 OR KIAA1453.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [2]
RP SEQUENCE OF 1-954 FROM N.A.
RA Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 494-1121 FROM N.A.
RC TISSUE=Testis;
RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SUBCELLULAR LOCATION.
RX MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
```

```
DR InterPro: IPR001394; Peptidase_C19.
DR Pfam: PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW ubi conjugation pathway; Hydrolase; Thiol protease; Multigene family;
KW Nuclear protein.
FT ACT_SITE 131 BY SIMILARITY.
FT ACT_SITE 373 BY SIMILARITY.
FT ACT_SITE 382 BY SIMILARITY.
FT CONFLICT 573 D -> G (IN REF. 2).
FT CONFLICT 806 R -> Q (IN REF. 3).
FT CONFLICT 828 C -> R (IN REF. 3).
FT CONFLICT 958 K -> KKK (IN REF. 3).
SQ SEQUENCE 1121 AA; 122626 MW; 2476F5128CBAB5A5 CRC64;

Query Match 1.8%; Score 14; DB 1; Length 1121;
Best Local Similarity 100.0%; Pred.No. 1.8e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 310 SNVLTLSLKEFANF 323
Db 321 SNVLTLSLKEFANF 334

Search completed: August 10, 2004, 20:15:42
Job time : 15 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:59:54 ; Search time 50 Seconds
(without alignments)
4862.086 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIYDKASESDPSAYQNP.....AAESLEEDPAASLFFPSEGG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	ID	Description
1	4028	99.0	1251	US-10-399-645-7	Sequence 7, Appli
2	4028	99.0	1316	US-10-379-981-9	Sequence 9, Appli
3	3399	98.3	762	US-10-163-547-1	Sequence 1, Appli
4	3198.5	78.6	1326	US-10-379-981-12	Sequence 12, Appli
5	1423.5	35.0	1123	US-09-972-525-1	Sequence 1, Appli
6	1150.5	28.3	565	US-10-371-905A-20	Sequence 20, Appli
7	1148	28.2	530	US-10-107-695B-2	Sequence 2, Appli
8	1142.5	28.1	530	US-10-371-905A-16	Sequence 16, Appli
9	1139.5	28.0	530	US-10-371-905A-14	Sequence 14, Appli
10	1135.5	27.9	530	US-10-371-905A-18	Sequence 18, Appli
11	1132.5	27.8	715	US-10-371-905A-22	Sequence 22, Appli
12	1128.5	27.7	1021	US-10-371-905A-8	Sequence 8, Appli
13	1126.5	27.7	633	US-10-371-905A-24	Sequence 24, Appli
14	1125	27.6	472	US-10-408-765A-2752	Sequence 2752, Ap
15					

16	1121.5	27.6	1016	12	US-10-371-905A-4	Sequence 4, Appli
17	1118.5	27.5	530	12	US-10-371-905A-10	Sequence 10, Appli
18	1114	27.4	398	12	US-10-371-905A-42	Sequence 42, Appli
19	1113.5	27.4	530	10	US-09-372-348-11	Sequence 11, Appli
20	1111.5	27.3	530	12	US-10-371-905A-6	Sequence 6, Appli
21	1096	26.9	530	12	US-10-371-905A-34	Sequence 34, Appli
22	1073	26.4	530	12	US-10-371-905A-26	Sequence 26, Appli
23	1052	25.8	530	12	US-10-371-905A-32	Sequence 32, Appli
24	1049.5	25.8	530	12	US-10-371-905A-36	Sequence 36, Appli
25	1037	25.3	357	12	US-10-371-905A-40	Sequence 40, Appli
26	1030	25.3	530	12	US-10-371-905A-30	Sequence 30, Appli
27	980.5	24.1	497	12	US-10-371-905A-28	Sequence 28, Appli
28	914	22.5	748	10	US-09-372-348-15	Sequence 15, Appli
29	898	21.8	545	10	US-09-372-348-17	Sequence 17, Appli
30	846.5	20.8	526	10	US-09-372-348-11	Sequence 16, Appli
31	799.5	19.6	314	15	US-10-274-639-11	Sequence 11, Appli
32	799.5	19.6	314	16	US-10-333-574-11	Sequence 11, Appli
33	787	19.3	844	16	US-10-437-963-132523	Sequence 132523, A
34	786.5	19.3	928	12	US-10-425-114-65150	Sequence 65150, A
35	760	18.7	457	15	US-10-369-493-22645	Sequence 22645, A
36	735.5	18.1	993	16	US-10-437-963-182980	Sequence 182980, A
37	728	17.9	549	12	US-10-424-599-233768	Sequence 233768, A
38	719	17.7	1003	12	US-10-425-114-47268	Sequence 47268, A
39	718	17.6	1204	16	US-10-437-963-161587	Sequence 161587, A
40	701	17.2	1010	16	US-10-437-963-131687	Sequence 131687, A
41	690	17.0	735	12	US-10-425-114-54385	Sequence 54385, A
42	666	16.4	919	16	US-10-437-963-183316	Sequence 183316, A
43	658.5	16.2	636	16	US-10-437-963-116674	Sequence 116674, A
44	657.5	16.2	563	12	US-10-425-114-48696	Sequence 48696, A
45	652.5	16.0	248	12	US-10-371-905A-46	Sequence 46, Appli

ALIGNMENTS

RESULT 1

US-10-399-645-7
; Sequence 7, Application US/10399645
; Publication No. US20040029249A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; LEE, Ernestine A.
; APPLICANT: HAFALTA, April J.A.; YUE, Henry
; APPLICANT: LAU, Freet G.; YAO, Monique G.
; APPLICANT: LU, Yan; CHAWLA, Narinder K.
; APPLICANT: WARREN, Bridget A.; LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.; DELEGEANE, Angelo M.
; APPLICANT: BURFORD, Neil; BOROWSKY, Mark L.
; APPLICANT: LEE, Sally; XU, Yuming
; APPLICANT: GRIFFIN, Jennifer A.; KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Yalda; ELLIOTT, Vicki S.
; APPLICANT: SWARNAKAR, Anita; RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Dannie B.; TRIBOULEY, Catherine M.
; APPLICANT: LO, Terence P.; AU-YOUNG, Janice K.
; APPLICANT: THANGAVELU, Kavitha; KEARNEY, Liam
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0263 USN
; CURRENT APPLICATION NUMBER: US/10399,645
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: PCT/US01/51034
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 60/241,573
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/243,643
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/245,256
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/248,395
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/249,826
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,303

Wed Aug 18 13:53:05 2004

```

; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,981
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040029249A1 7997881CD1
US-10-399-645-7

Query Match          99.0%; Score 4028; DB 12; Length 1251;
Best Local Similarity 99.7%; Pred. No. 5.5e-286;
Matches 767; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
Db 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFCMCTQWQAHITQALSNPGDVVKPMFVI 180
Db 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFCMCTQWQAHITQALSNPGDVVKPMFVI 180
QY 181 NEMRRITARHLRFQNGEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
Db 181 NEMRRITARHLRFQNGEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
QY 181 NEMRRITARHLRFQNGEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
Db 181 NEMRRITARHLRFQNGEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
Db 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
Db 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQMDNSIVSTSDIRSVLSQQAAYVLYIRSHDVKNNGE 420
Db 361 VHTGFNCHAGHYFCYIKASNGLWYQMDNSIVSTSDIRSVLSQQAAYVLYIRSHDVKNNGE 420
QY 421 LTHPTSPGSSPRPVISQSVVTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480
Db 421 LTHPTSPGSSPRPVISQSVVTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480
QY 481 MSSPNGNSVNRASPVNASAVQNSVNRSSVIEHPKPKQKITIISIHKLPRVQCQSQPN 540
Db 481 MSSPNGNSVNRASPVNASAVQNSVNRSSVIEHPKPKQKITIISIHKLPRVQCQSQPN 540
QY 541 LHSNSLENPTKVPSSITITNSAVOSTSNASTMSVSSVKTKPIPRSESCSQPMNGKSKLN 600
Db 541 LHSNSLENPTKVPSSITITNSAVOSTSNASTMSVSSVKTKPIPRSESCSQPMNGKSKLN 600
QY 541 LHSNSLENPTKVPSSITITNSAVOSTSNASTMSVSSVKTKPIPRSESCSQPMNGKSKLN 600
Db 541 LHSNSLENPTKVPSSITITNSAVOSTSNASTMSVSSVKTKPIPRSESCSQPMNGKSKLN 600
QY 601 SSVLPVYGAESSESDDESKGLGKENGIGTIVSSHSPQDQADEEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSESDDESKGLGKENGIGTIVSSHSPQDQADEEATPHELQEPMTLNGA 660
QY 601 SSVLPVYGAESSESDDESKGLGKENGIGTIVSSHSPQDQADEEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSESDDESKGLGKENGIGTIVSSHSPQDQADEEATPHELQEPMTLNGA 660
QY 661 NSADSDDPKENGLAPDGASCOQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720
Db 661 NSADSDDPKENGLAPDGASCOQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAERGPPEDRDAPGSPAESLSEPDAAAAL 769
Db 721 RLSNKLKSTDEMSAPGAERGPPEDRDAPGSPAESLSEPDAAAAL 769
QY 721 RLSNKLKSTDEMSAPGAERGPPEDRDAPGSPAESLSEPDAAAAL 769
Db 721 RLSNKLKSTDEMSAPGAERGPPEDRDAPGSPAESLSEPDAAAAL 769

RESULT 2
US-10-379-981-9

```

```

; Sequence 9, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; PRIOR FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-981-9

```

```

Query Match          99.3%; Score 4028; DB 15; Length 1316;
Best Local Similarity 99.7%; Pred. No. 5.9e-286;
Matches 767; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
Db 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFCMCTQWQAHITQALSNPGDVVKPMFVI 180
Db 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFCMCTQWQAHITQALSNPGDVVKPMFVI 180
QY 181 NEMRRITARHLRFQNGEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
Db 181 NEMRRITARHLRFQNGEDAHEFLQYTVDMQKACLNGSNKLDRTHTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
Db 241 RVKCLNCKGVSDTFDPLVDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
Db 301 SKRFTIHRSSNVLTLSLKRFPANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQMDNSIVSTSDIRSVLSQQAAYVLYIRSHDVKNNGE 420
Db 361 VHTGFNCHAGHYFCYIKASNGLWYQMDNSIVSTSDIRSVLSQQAAYVLYIRSHDVKNNGE 420
QY 421 LTHPTSPGSSPRPVISQSVVTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480
Db 421 LTHPTSPGSSPRPVISQSVVTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480
QY 481 MSSPNGNSVNRASPVNASAVQNSVNRSSVIEHPKPKQKITIISIHKLPRVQCQSQPN 540
Db 481 MSSPNGNSVNRASPVNASAVQNSVNRSSVIEHPKPKQKITIISIHKLPRVQCQSQPN 540
QY 541 LHSNSLENPTKVPSSITITNSAVOSTSNASTMSVSSVKTKPIPRSESCSQPMNGKSKLN 600
Db 541 LHSNSLENPTKVPSSITITNSAVOSTSNASTMSVSSVKTKPIPRSESCSQPMNGKSKLN 600
QY 601 SSVLPVYGAESSESDDESKGLGKENGIGTIVSSHSPQDQADEEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSESDDESKGLGKENGIGTIVSSHSPQDQADEEATPHELQEPMTLNGA 660
QY 661 NSADSDDPKENGLAPDGASCOQPALHSENPFPAKANGLPGLMPAPLLSLPEDKILETF 720

```


Db 661 NSADSDSPKENGAPDAGSCQGFALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAASL 769
Db 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAAGL 769
RESULT 3
US-10-163-547-1
; Sequence 1, Application US/10163547
; Publication No. US200303037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Myers, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US200303037350A1el nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-Ts Metalloprotease and uses
; TITLE OF INVENTION: therefore
; FILE REFERENCE: 35800/247836(5800-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,408
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 09/496,005
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-547-1
Query Match 98.38; Score 3999; DB 14; Length 762;
Best Local Similarity 99.94; Pred. No. 3.5e-284;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTIVDKASESDPSAYONQPGSEAVSPGMDAGSAGWAVSSINDVSNHTLSLGPVPGA 60
Db 1 MTIVDKASESDPSAYONQPGSEAVSPGMDAGSAGWAVSSINDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPFLANYMLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180
Db 121 FANAALQCLTYTTPFLANYMLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180
QY 181 NEMRIARLRFNGQDAHEFLQYTVDMQKACLNGSNKLDRTQATTLVCQIFGGYLR 240
Db 181 NEMRIARLRFNGQDAHEFLQYTVDMQKACLNGSNKLDRTQATTLVCQIFGGYLR 240
QY 241 RVKCLNCKGVSDTFDPYLDITLIEKAAQSVNKALEQFVKPEQLDGENSYKSKCKMWPA 300

Db 241 RVKCLNCKGVSDTFDPYLDITLIEKAAQSVNKALEQFVKPEQLDGENSYKSKCKMWPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFANFTGCKIAKVDKYPEYLDIRPYMGPONGEPYVYLYAVL 360
Db 301 SKRFTIHRSSNVLTLSLKRFANFTGCKIAKVDKYPEYLDIRPYMGPONGEPYVYLYAVL 360
QY 361 VHTGENCHAGHYFCYIKASNGLIWQNDISVSTDIRSVLSQOAYVLFYIRSHDVKNNGE 420
Db 361 VHTGENCHAGHYFCYIKASNGLIWQNDISVSTDIRSVLSQOAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTHSPGQSSPRPVISQRVVTNNKQAPGFTGPOLPSHMIKNPHEHNGTGLKDTSPSS 480
Db 421 LTHPTHSPGQSSPRPVISQRVVTNNKQAPGFTGPOLPSHMIKNPHEHNGTGLKDTSPSS 480
QY 481 MSSPNGNSSVNRASFPVNASAVQNVSNRSSVPEHPKKQKITTISIHNKLPVRQCSQSPN 540
Db 481 MSSPNGNSSVNRASFPVNASAVQNVSNRSSVPEHPKKQKITTISIHNKLPVRQCSQSPN 540
QY 541 LHSNSLENPTKPVFSSITINSAVOSTSNASTMSVSSKVTKIPRSESCSQPVMNGSKLIN 600
Db 541 LHSNSLENPTKPVFSSITINSAVOSTSNASTMSVSSKVTKIPRSESCSQPVMNGSKLIN 600
QY 601 SSVLPVYGAESSESDSDESKGLGKENGIGTIVSSHSPGQDAEDEATPHELOEPMTLNGA 660
Db 601 SSVLPVYGAESSESDSDESKGLGKENGIGTIVSSHSPGQDAEDEATPHELOEPMTLNGA 660
QY 661 NSADSDSPKENGAPDAGSCQGFALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
Db 661 NSADSDSPKENGAPDAGSCQGFALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAASL 762
Db 721 RLSNKLKSTDEMSAPGAEPPEDRDAPQPSGAASLEPPDAAGL 762
RESULT 4
US-10-379-981-12
; Sequence 12, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE OF INVENTION: MURINE ORTHOLOG
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1326
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-379-981-12
Query Match 78.6%; Score 3198.5; DB 15; Length 1326;
Best Local Similarity 80.14%; Pred. No. 3.9e-225;
Matches 620; Conservative 50; Mismatches 87; Indels 17; Gaps 7;
QY 1 MTIVDKASESDPSAYONQPGSEAVSPGMDAGSAGWAVSSINDVSNHTLSLGPVPGA 60
Db 1 MTIVDK-TFSPDSTCQNPQGSCEAVSPEDMTDGSAGWAVSSISDVSSHPLGVPVGA 59
QY 61 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 60 VVYSSSVDPKSPKQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 119
QY 121 FANAALQCLTYTTPFLANYMLSHEHSKTCCHAEFCMCTMQAHTQALSNPDGVIKPMFVI 180

Db 120 FANAALQCLTYTPPLANMLSHEHSKTCBAEGFCMCTMQTHIQAALNSPQGVIVKMEVI 179
QY 181 NEMRIARHLRFQNGQEDAEHFLQYTVDAKACALNGSNKLDHRTQATTLVQCIFFGYLRS 240
Db 180 NEMRIARHLRFQNGQEDAEHFLQYTVDAKACALNGSNKLDHRTQATTLVQCIFFGYLRS 239
QY 241 RVKCLNCKGVSDTDPYLDITLTKAAQSVNKALEQFVKPQOLGDSYKSKCKKQVPA 300
Db 240 RVKCLNCKGVSDTDPYLDITLTKAAQSVNKALEQFVKPQOLGDSYKSKCKKQVPA 299
QY 301 SKRFTIHRSSNVLTSLKRFANFTGCKTAKDVKYPEYLDIRPYMSQNGEPTVVLYAVL 360
Db 300 SKRFTIHRSSNVLTSLKRFANFTGCKTAKDVKYPEYLDIRPYMSQNGEPTVVLYAVL 359
QY 361 VHTGNCCHAGHYFCYIKASNGLIWQMNDSIVSTDSIRSVLSQAAVLYFIRSHDVKNNGE 420
Db 360 VHTGNCCHAGHYFCYIKASNGLIWQMNDSIVSTDSIRAVLNQAAVLYFIRSHDVKNNGE 419
QY 421 LTHPTSPQSSPRVIVSORVVTNKQAAPGFTGPQLPSHMIKPNPHLNGTGPLKDTTPSS 480
Db 420 SAHPAHSPQSSPRVIVSORVVTNKQAAPGFTGPQLPSHVMKNTPHLNGTTPVKDTPSS 479
QY 481 MSSPNCNSVNRASPVNASASVQNSVNRSSVIVPEHPKKOKITISIHNNKLPVRQCSQBN 540
Db 480 VSSPNCNTSVNRASPVNASASVQNSVNRSSVIVPEHPKKOKITISIHNNKLPVRQCSQBN 539
QY 541 --LHNSLENPKPVPSSTITN-SAVQSTSNASTMSVSSKVTPIPRSSCSQPVNNGKS 597
Db 540 NSLHGPCLEAPSPSTITNPSAIQSTSNVPTTSTP-----SEACPKPVNNGKA 591
QY 598 KLNSSVLPVYGAESSEDSDEESKGLKENGIGTIVSSH--PGQDAED-EAATPHELOEP 654
Db 592 KVGASVLPVYGAESSEDSDEESKGLAKENGVDMMAGTHSDRPEAAADDGAASSHLOEP 651
QY 655 MTLGANSADSDPKENGLAPDGASQCGOPALHSENPFAKANGLPKMLPAPLLSLPED 714
Db 652 VLLGANSADSDS--QENSLAFDSASCQOPELHTENLFKLNGLPKCTVPAPLQSPVED 709
QY 715 KILETFLRSLNKLKSTDEMSAFGAERPPEDRDAEPQGSAALESLEBPDAAS 768
Db 710 RILETFLRSLNKLKSTDEMSAFGAERPPEDRDAEPQGSAALESLEBPDAAS 763

RESULT 5
US-09-972-525-1
; Sequence 1, Application US/09972525
; Publication No. US20020086407A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/972,525
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/408,865
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-525-1
Query Match 35.0%; Score 1423.5; DB 12; Length 1123;
Best Local Similarity 37.3%; Pred. No. 4.le-95;
Matches 347; Conservative 112; Mismatches 238; Indels 233; Gaps 25;
QY 1 MTLVDKASSDSPAYQNGQSSAEVSPGDMAGSAGVASSINDV----- 47
Db 1 MPVLDKLKEAL-----KPKRGASDDGELGKLILASSAKKVLQKIEFEFPASKSFYQL 53
QY 48 ----SNHTLSLGPVPGAVVY-SSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLK 102

Db 54 EALKSKYVLLNPKTEGASRHSKGGDPPARRQSGSEHTYSCGDPVAPQKVLFPPTLSLR 113
QY 103 WOQTHRVGAGLONLGNCTCFANAALQCLTYTPPLANMLSHEHSKTCBAEGFCMCTMOAH 162
Db 114 WERVFRVAGLHNLNCTCFNLNATIQCLTYTPPLANMLSHEHSKTCBAEGFCMCTMOAH 173
QY 163 ITOALSNGPDVVKMFVINEMRRIRARHLRFQNGQEDAEHFLQYTVDAKACALNGSNKLDH 222
Db 174 IVQAFANSNAIKPVFIRDILKKIARHFRFQNGQEDAEHFLQYTVDAKACALNGCAKLD 233
QY 223 HTQATTLVQCIFFGYIRSRVKLCKNGVSDTDPYLDITLTKAAQSVNKALEQFVKPEQ 282
Db 234 QTQATTLVQCIFFGYIRSRVKCSVCKSVSDTDPYLDVLALEIRQAANIVRALELVKADV 293
QY 283 LDGENSYKSKCKKQVPAKSKFTIHRSSNVLTSLKRFANFTGCKTAKDVKYPEYLDIRP 342
Db 294 LSGENAYNCAKCKKVPASKKFTIHRSSNVLTSLKRFANFTGCKTAKDVKYPEYLDIRP 353
QY 343 YMSQNGEPIVVVLYAVLVHTGFCNCHAGHYFCYIKASNGLIWQMNDSIVSTDSIRSVLSQ 402
Db 354 YMSQNGEPIVVVLYAVLVHTGFCNCHAGHYFCYIKASNGLIWQMNDSIVSTDSIRSVLSQ 413
QY 403 QAYVLYFIRSHDVKNNGE---LTHPTSPQSSPRP-----VLSQVVTNKQ-- 446
Db 414 QAYVLYFIRSHDVKNNGE---LTHPTSPQSSPRP-----VLSQVVTNKQ-- 473
QY 447 -----AAGFTGPOLPS-----HMIKPNPHLNGTGPLKD 475
Db 474 GTMKKPHTEIEGVP--SRNGSTLGLKQNGCTPPLKPSGSPKLSQVTHM---PTILD 530
QY 476 TPSSSMSSP-----NGNSS-----VNRASPV----- 496
Db 531 DFGKVKKPPAPPOHPSPTAQGLPGTSNSNSRSGSQROGSDRDVVLSTSPKILATAT 590
QY 497 -----NASAVQNSVNRSSVIVPEHPK-----KOKITISI--H 527
Db 591 ANGHGLKGNDESAGLDRRGSSSSPEHSASDSSTKAPQTPRSGAAHLCDSQETNCSTAGH 650
QY 528 NKLPVRQCSQ-----PNLHNSLE-----NPT 550
Db 651 SKTPPSGADSKTVKLKSPVLSNTTETPASTMGPAPPAKKALSAKASTLWRATGNDRPP 710
QY 551 KPVPSTITNSAVQSTSN---ASTMSV-SKVTKPIPRSESCSQPVYMGKSKLNSVLPV 606
Db 711 PPSFSSDLTHP--MKTSHPVVASTWVHRARAVSPAPQSSSRLOPPSPHPTLLSSTPKP 768
QY 607 YGAESSEDSDEESKGLKENGIGTIVSSHSPGQDAEDEATPHELOEPMTLNGANSADSD 666
Db 769 PGTSEPRSCSSISTALPQVN-----EDLVSLPHOLPEA----- 801
QY 667 SDPKENGLAPDGASQCGOP-ALHSENPFAKANGLPKMLPAPLLSPEDKILETFLRSLN 725
Db 802 SEPPQSPSEKRRKKTIVGEFQRLGSE-----TFLPOHIREA--TAAPHGK-----RKRK 847
QY 726 LKGSDEMSAPGAERPPEDRDAEPQGSF 755
Db 848 KKKRPEDTAASALQEG-----QTORQPGSP 872

RESULT 6
US-10-371-905A-20
; Sequence 20, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: US2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent in version 3.1

; SEQ ID NO 20
 ; LENGTH: 565
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC FEATURE
 ; OTHER INFORMATION: human DUB4.8
 US-10-371-905A-20

Query Match 28.3%; Score 1150.5; DB 12; Length 565;
 Best Local Similarity 46.7%; Pred. No. 1.6e-75;

Matches 244; Conservative 78; Mismatches 162; Indels 39; Gaps 8;

QY	28	PCDMDAGSASGAVSSLDVSNHTLSLGPVPGAVVYSSSVDPKSPQKDQALGDGIA	87
DB	33	PVDMEDDSLVLGGGEWFHFSKLT--SSRPDAFAFIEIQTSLPEKSPSCETRVLDLDDIA	91
QY	88	PPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHEHST	147
DB	92	FVARQLAPREKLPPLSSRPAAVAGAGLQNMGNCTCYVNASLQCLTYTPPLANYMLSREHSQT	151
QY	148	CHABGFCMCTMQAHITCALSNPGDVIKPMFVINEMRIARHLRFGNOEDAHEFLQYTVTD	207
DB	152	CHRHGCMCTMQAHITRALNPGHVIQP-----SQALAAAGFHRGKQEDAEHFLMFTVD	205
QY	208	AMQKACLNGSKLDRHTQATTLVCOIFGGYLSRVKCLNCKGVSDTFDPYLDITLEIKAA	267
DB	206	AMKACLPQKQVHDHSHKDTTLHQIFGGYWSQIKCLHCHGISDTPYLDIALDIOAA	265
QY	268	OSVNAKALROFVKPQDGENSYKCKCKMVPASKRFTIHRSSNVLTSLKRFANFTGGK	327
DB	266	OSVQOALEQLVKPELNGENAYHCGVCLQAPASKTTLHTSAKVLIIVLKRFSVDVTGNK	325
QY	328	IAKDVKYPEYLDIRPYMSQNPGEPIVYVLYAVLVHTGFCHAGHYFCYIKASNGLWQMN	387
DB	326	IAKNVQPECLDMQPYMSQQNTGELVYVLYAVLVHAGWSCHNGHYFSYVKAQEGQWYKMD	385
QY	388	DSIVSTSDIRSVLSQQAYVLYFYIR-----SHDVKNKGELTHTSPGSSPRPVISQR	440
DB	386	DAEVTAAISITSVLSQQAYVLYFYIQKSEWHSVSRSRGRE-----PRALGAED	433
QY	441	VVTKQAAPGFI-----GPQLPSHMIKNPPLHNGTGPLKOTPPSSMSPNNGN-SSVN	491
DB	434	--TDRRATQELKRDHPCLQAPDELHVERATQESTLDHWKFLQEQNKTKPENVKVE	491
QY	492	RASPVNASASVQNVSNRNVSSVIEHPKXKQKITISIHKNKLPVRQ	534
DB	492	GTLP--PDVLVIHQSKYKCGMKNHHPQEQSSLLMSSTPTHQ	532

RESULT 7

US-10-107-695B-2
 ; Sequence 2, Application US/10107695B
 ; Publication No. US2003002201A1
 ; GENERAL INFORMATION:

; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 68999, Human Ubiquitin
 ; TITLE OF INVENTION: Carboxyl-Terminal Hydrolase Family Member and Uses Therefor
 ; FILE REFERENCE: MPI01-076P/RCPIW
 ; CURRENT APPLICATION NUMBER: US/10/107,695B
 ; CURRENT FILING DATE: 2002-09-10
 ; PRIOR APPLICATION NUMBER: 60/279,184
 ; PRIOR FILING DATE: 2001-03-27
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 2

LENGTH: 530

TYPE: PRT

ORGANISM: Homo sapiens

US-10-107-695B-2

Query Match

Best Local Similarity 28.2%; Score 1148; DB 14; Length 530;

Matches 243; Conservative 77; Mismatches 161; Indels 39; Gaps 8;

Matches 252; Conservative 89; Mismatches 164; Indels 80; Gaps 12;

QY	31	MDAGSASGAVSSLDVSNHTLSLGPVPGAVVYSSSVDPKSPQKDQALGDGIA	90
DB	1	MEDDSLVLGGGEWFHFSKLT--SSRPDAFAFIEIQTSLPEKSPSCETRVLDLDDIA	91
QY	91	KVLPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHEHST	150
DB	60	QLAPREKLPPLSSRPAAVAGAGLQNMGNCTCYVNASLQCLTYTPPLANYMLSREHSQT	119
QY	151	EGFCMCTMQAHITCALSNPGDVIKPMFVINEMRIARHLRFGNOEDAHEFLQYTVDA	210
DB	120	HKGCMCTMQAHITRALNPGHVIQP-----SQALAAAGFHRGKQEDAEHFLMFTVD	173
QY	211	KACLNGSKLDRHTQATTLVCOIFGGYLSRVKCLNCKGVSDTFDPYLDITLEIKAA	270
DB	174	KACLPQKQVHDHSHKDTTLHQIFGGYWSQIKCLHCHGISDTPYLDIALDIOAA	233
QY	271	NKALEOFVKPQDGENSYKCKCKMVPASKRFTIHRSSNVLTSLKRFANFTGGK	330
DB	234	QOALEQLVKPELNGENAYHCGVCLQAPASKTTLHTSAKVLIIVLKRFSVDVTGNK	293
QY	331	DVKYPEYLDIRPYMSQNPGEPIVYVLYAVLVHTGFCHAGHYFCYIKASNGLWQMN	390
DB	294	NVQPECLDMQPYMSQQNTGELVYVLYAVLVHAGWSCHNGHYFSYVKAQEGQWYKMD	353
QY	391	VSTSDIRSVLSQQAYVLYFYIR-----SHDVKNKGELTHTSPGSSPRPVISQR	443
DB	354	VTAAISITSVLSQQAYVLYFYIQKSEWHSVSRSRGRE-----PRALGAED--T	399
QY	444	NKQAAPGFI-----GPQLPSHMIKNPPLHNGTGPLKOTPPSSMSPNNGN-SSVN	495
DB	400	DRRATQELKRDHPCLQAPDELHVERATQESTLDHWKFLQEQNKTKP	448
QY	496	VNASASVQNVSNRNVSSVIEHPKXKQKITISIHKNKLPVRQ	549
DB	449	-----EFNVKVEGTLPDVLVIHQSKYKCGMKNHHP-----EQSSLLMSSTNP	494
QY	550	TKVPVSTITNSAVQSTNSATM-SVSSKVTKPIRSESCSQPVW	593
DB	495	TD-----QESMNTGTLASLQGRTRRAKGNKHKCRSIL	527

RESULT 8

US-10-371-905A-16
 ; Sequence 16, Application US/10371905A
 ; Publication No. US20030224969A1
 ; GENERAL INFORMATION:

; APPLICANT: Hahn, Chang
 ; APPLICANT: Liu, Hong
 ; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
 ; FILE REFERENCE: USAV2001/0185
 ; CURRENT APPLICATION NUMBER: US/10/371,905A
 ; CURRENT FILING DATE: 2003-02-20
 ; NUMBER OF SEQ ID NOS: 54
 ; SOFTWARE: PatentIn version 3.1

SEQ ID NO 16

LENGTH: 530

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: MISC FEATURE

OTHER INFORMATION: human DUB4.6

US-10-371-905A-16

Query Match

Best Local Similarity 28.1%; Score 1142.5; DB 12; Length 530;

Matches 243; Conservative 77; Mismatches 161; Indels 39; Gaps 8;

QY 31

MDAGSASGAVSSLDVSNHTLSLGPVPGAVVYSSSVDPKSPQKDQALGDGIA
 1 MEDDSLVLGGGEWFHFSKLT--SSRPDAFAFIEIQTSLPEKSPSCETRVLDLDDIA

DB 1

Query Match	28.0%;	Score 1140.5;	DB 12;	Length 530;
Best Local Similarity	46.5%;	Pred. No. 7.7e-75;		
Matches 242;	Conservative 78;	Mismatches 161;	Indels 39;	Gaps 8;
QY	31	MDAGSASGAVSSLNDVSNHTLSLCPVEGAVYSSSSVPDKSKSPQKDQALGDIAPPQ	90	
DB	1	MEDDSLILGGEQWFNHFSEKLT--SSRPDAAFAEIQTSLPEKSPSLCETRVLDLDDLPVA	59	
QY	91	KVLPFSEKICLKWOOTHVAGLQNLGTCFANALQCLTTPPLANYMLSHESKTC	150	
DB	60	QLAPREKLPLUSSRRPAAVAGLQNMGNITCVNNSLQCLTTPTPPLANYMLSHESQTC	119	
QY	151	EGFCNMCTMQAHIQTALSNPGSVIKPMFVINEMRRIARHLAFGNQDAHEFLQYTV	210	
DB	120	HKGCMCLMTQAHITRALNHPGHVIOF-----SQALAAGFHTRGKQDAHEFLMF	173	
QY	211	KACLGNSKNLDRHTQATTLVCQIFGGYLRSRVKLCNCKGVSDTFPFDYLDITLEIKAAQSV	270	

```

; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human J04.5
US-10-371-905A-14

      Query Match      28.3%;      Score 1139.5;      DB 12;      Length 574;
      Best Local Similarity 46.58;      Pred. No. 1e-74;
      Matches 243;      Conservative 164;      Mismatches 164;      Indels 39;      Gaps 8;

QY      28      PGDMDAASGASMGVSSLDVSNHTLSLGPVGEAVVYSSSVDPKSPQKQALGDGIA 87
DB      33      PVDMEEDSLYLGGEQWFNFHFSKLT-SSRPDAAFAEIQTISLPEKSPSCTETRVLDCCDLA 91

QY      88      PPQKVLPPSEKICLKWOOTHRVAGLQNLGNTCFANAALQCLTYTTPPLANMYLMSHSKT 147
DB      92      PVARQLAPREKPLPSNRPAAGVAGLQMGNTCTVNASLQCLTYTTPPLANMYLMSHSQT 151

QY      148      CHAFGFCMCTMQAHIICLALSNPQGVITKPMFVINEMERRIARHLRFQGNODEAHFELQYTVD 207
DB      152      CHRHKGCMLCTMQAHIITRALHNPGHVITQP-----SQALAGFIRGQEDAEHFLMFIVD 205

QY      208      AMOKACLGNSKLDHRHTQATTLVCOIFGGYLSRVKCLNCKGVSDTFDPYLDITLEIKAA 267
DB      206      AMKRACLPGHKQVDHHSKDTLLIHQIFGGYWSQIKCLCHCHGISDTFDPYLDIALDIQAA 265

QY      268      QSVNKALBQFVKPCELDGENSYKSKCKKQMPAPAKRFTIHRSSNVLTLSLKRPFANFTGK 327
DB      266      QSVQALBQLKPEBELNGENAYHCGVCLQRAPASKTLLTHTSAKVLILVLRFSDTVGNK 325

QY      328      TAKDVKYPEYLDIREYNSQNPGEPIVVYLVAVLHTGNCCHAGHYFCVYKASNGLWQWN 387
DB      326      IDKNVQYCECLDMKLYMSQTSNGPLVYLVAVLVHAGWSCHNGHVFYSVKVQEGQYKWD 385

```



```
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1021
; TYPE: PRP
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.2a
US-10-371-905A-8

Query Match      27.7%; Score 1128.5; DB 12; Length 1021;
Best Local Similarity 46.1%; Pred. No. 1.5e-73;
Matches 241; Conservative 78; Mismatches 165; Indels 39; Gaps 8;

QY      28  PCDMDAGSAGWAVSSLDVNSHTLSLGPVPGAVVYSSSSVDPKSPKQDQALGDGIA 87
DB      489  PVDMDSDSLYLGGWQFNHFSKLT--SSRPDAFAEIQTSLSEKPLSCETRVDLCDLLA 547

QY      88  PPQKVLFFSEKICLKQWQTHRVGAGLNGTCTCFANAALQCLTYTTPPLANYMLSHEHST 147
DB      548  FVARQLAPREKPLPLSSRRPAAVAGAGLQNGNTCYVNASLQCLTYTTPPLANYMLSREHST 607

QY      148  CHAEGFCMMCTQAHITQALSNPQDVIKPMFVINEMRRITARHLRFGNQDAHEFLQYTV 207
DB      608  CHRHGCMCTMQAHITRALHNPCHVIOQ-----SQALAAAGFHRGQDAHEFLMFTVD 661

QY      208  AMOKACLSGNKLDHRTQATTLVQIFGGYLSRVKCLCKGVSDFDPYLDITLLEIKAA 267
DB      662  AMKACLPQGHQVDDHHSKDTLLHQIFGGYWSQIKLCHGIGSDFDPYLDIALDIQAA 721

QY      268  QSNKALEQVFKPEQLDGENSKYKCKKQKVPASKRFTTHRSSNVLTLSLKRFANFTGK 327
DB      722  QSVQALEQLVXPEELNGENAYHSGVCLQRAPASKTLTLTSAKVLILVLFKRFSDVTGNK 781

QY      328  IAKVVKPYEYLDIRPMSQNGEPIYVLYAVLVHTGFNCHAGHYCYIKASNLWYQWN 387
DB      782  IAKNVQYPECLDMQYMSQQNTGLVYLYAVLVHAGWSCHNGHYFSYKAOBQGWYKMD 841

QY      388  DSTVSTDIRSVLSQAYVLYFYIR-----SHDVKNKGELTHPTSPGQSSPRPVISOR 440
DB      842  DAEVTAASITSAUSQAYVLYFYIQSEWERHSESVSRGRE-----PRALGTED 889

QY      441  VVTNKQAAPGFI-----GPQLPSHMIKNPPLHNGTGLKDTPTSSMSSPNGN--SSVN 491
DB      890  --TDREATQELKRDHPCLQAPDELDEHLVERATQESTLDHWKFLQEQNKTKPFNVKVE 947

QY      492  RASPVNASASVQWNSVRSSVIEHPKQKQITISIHKNLPVRQ 534
DB      948  GTLP--PDVLVHQSRYKCGMKNHHPQSSLLNLSSTPTHQ 988

RESULT 14
US-10-371-905A-24
; Sequence 24, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
```

```
; SEQ ID NO 24
; LENGTH: 633
; TYPE: PRP
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human LUB4.11
US-10-371-905A-24

Query Match      27.7%; Score 1126.5; DB 12; Length 633;
Best Local Similarity 44.6%; Pred. No. 1.1e-73;
Matches 246; Conservative 84; Mismatches 169; Indels 53; Gaps 11;

QY      1  MTIV--DKASESSDPEAYQNPQGSSEAVSPGMDAGSASWGAVSGLNDVSNHTLSLGPVP 58
DB      84  LTLIRSESALSSSCFVAERMAAAKI---DMEDDSLVLGGWQFNHFSKLT--SSRPDA 138

QY      59  GAVVYSSSSVDPKSPKQDQALGDGIAPQKVLFPSEKICLKQWQTHRVGAGLONLON 118
DB      139  APAEIQTSLSEKIS-----PLSYDLAPVARQLAPREKPLPLSSRRPAAVAGLQNGN 190

QY      119  TCFANAALQCLTYTTPPLANYMLSHEHSTKCHAEFGCMCTQAHITQALSNPQDVIKPMF 178
DB      191  TCYVNASLQCLTYTTPPLANYMLSREHSTQCHRHKGCMCTQAHITRALHNPCHVIOQ-- 248

QY      179  VINEMRRITARHLRFGNQDAHEFLQYTVDMQKACLSGNKLDHRTQATTLVQIFGGYL 238
DB      249  -----SQALAAAGFHRGQDAHEFLMFTVDAMKACLPKHKQVDHHSKDTLLHQIFGGY 304

QY      239  RSRVKCLCKGVSDFDPYLDITLLEIKAAQSVNKALEQVFKPEQLDGENSKYKCKKQMV 298
DB      305  RSQIKLCHGIGSDFDPYLDIALDIQAAQSVQQALEQLVKPEELNGENAYHSGVCLQRA 364

QY      299  PASKRFTTHRSSNVLTLSLKRFANFTGKIAKDVKYPEYLDIRPMSQNGEPIYVLYA 358
DB      365  PASKTLTLTSAKVLILVLFKRFSDVTGNKIAKNVQYPECLDMQYMSQQNTGLVYLYA 424

QY      359  VLVHVFNCHAGHYCYIKASNLWYQWMDSTVSTDIRSVLSQAYVLYFYIR-----411
DB      425  VLVHAEWSCHNGHYFSYKAOBQWYKMDAEVTAASITSVLSQAYVLYFYIQSEWERH 484

QY      412  SHDVKNKGELTHPTSPGQSSPRPVISORVTVNKQAAPGFI-----GPQLPSHMIKN 463
DB      485  SESVSRGRE-----PRALGAED--TDREATQELKRDHPCLQAPDELDEHLVER 530

QY      464  PHLNGTGLKDTPTSSMSSPNGN--SSVNRAFPVNASASVQWNSVRSSVIEHPKQKQI 522
DB      531  ATQESTLDHWKFLQEQNKTKPFNVKVKGTLP--PDVLVHQSRYKCGMKNHHPQSS 588

QY      523  TSIHKNLPVRQ 534
DB      589  LNLSSSTPTHQ 600

RESULT 15
US-10-408-765A-2752
; Sequence 2752, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Robin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenna, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 2752
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2752

Query Match      27.6%; Score 1125; DB 16; Length 472;
Best Local Similarity 50.9%; Pred. No. 8.9e-74;
Matches 239; Conservative 63; Mismatches 126; Indels 42; Gaps 9;

Qy 49 NHTLSL-GPVGAVV--YSSSSVPDKSPKQKQALGDGIAPPQKVLFPSEKICLKWQQ 105
Db 15 NHFSKLTSPRDAAFAETQRTSLPEKSPSCETRVLDLCYLAQVAPVQALAPREKLPSSRR 74
Qy 106 THRYGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHEHSKTCBAEGFCMMCTQAHITQ 165
Db 75 PAAVGAGLQNGNTCYVNASLQCLTYTPPLANYMLSREHSQTCHRHKGCMCTQAHITR 134
Qy 166 ALSNPGDVIKPMFVINEMRRIARHLRFGNQEDAEHEFLQYTVDMOKACLNGSNKLDRTQ 225
Db 135 ALHNPCHVIQF-----SQALAGFHRGKQEDAEHEFLMTVDAMKACLPGHKQVDHHSK 188
Qy 226 ATTIVCOIFGGLRSRVKCLNCKGVSTDFDYLITLTKAAQSVNKALEQFVKPEQLDG 285
Db 189 DTLIHQIFGGYWSQIKCLHCHGISDTDFDYLIDLDIAAQSQQALEQLVKPEELNG 248
Qy 286 ENSYKCSKCKMVPASKRFTTHRSSNVLTLSLKRFANFTGKIAKADVKYPEYLDIRPYS 345
Db 249 ENAYHCGVCLQRPASKTLTLHTSAKVLILVLRKFSQVDTGNKIANKVQYPECLDMQPYMS 308
Qy 346 QPNGEPIVYVLYAVLVHTGFNCHAGHYFCYIKASGLWYQWNSIVSTSDIRSVLSQAY 405
Db 309 QQNTGPLVYVLYAVLVHAGWSCHNGHYFVYKAEQGWYKIDDAEVTASSITSVLTQAY 368
Qy 406 VLFYIRSHDVANGEL-THPTSPGQSSPRPVI-----SQRVVTKQ-----AA 448
Db 369 VLFYIQSEWERHKSQVETVNP---SVRLSLFLFTQNSQFFLTQKQISLMEKKFSISG 425
Qy 449 PGFIGPQLPSHMIKNPPLNGTGPLK-----DTPSSSMSSPNGNSSVNR 492
Db 426 PAQARP-----SNPCHLSLGLGQRRGCGSPGSGTGRGPEGLSLRR 468
```

Search completed: August 10, 2004, 20:05:45
Job time : 53 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:13:12 ; Search time 21 Seconds
(without alignments)
3549.925 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 775
Sequence: 1 MTIVDKASESSDPAYQNQP.....AAESLEEPDAAASLFFPSEGG 775

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 10

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR 78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	1.3	887	2 C86453	CDS protein F9L11.

ALIGNMENTS

RESULT 1
C86453
CDS protein F9L11.5 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 25-Aug-2003
C;Accession: C86453
R;Theologus, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: C86453
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-887 <STO>
A;Cross-references: GB:AE005172; NID:96910582; PIDN:AAF31287.1; GSPDB:GN00141
C;Genetics:

A;Gene: F9L11.5
A;Map position: 1
C;Superfamily: ubiquitin C-terminal hydrolase Ubpl2

Query Match 1.3%; Score 10; DB 2; Length 887;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLQNLGNTCF 121
|||
Db 297 GLQNLGNTCF 306

Search completed: August 10, 2004, 20:17:08
Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:16:43 ; Search time 50 Seconds
(without alignments)

4862.086 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 775

Sequence: 1 MTIVDKASSESDPSAYQNQP.....AAESLEPDAASLFPFSEG 775

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1291235 seqs, 313682936 residues

Word size : 10

Total number of hits satisfying chosen parameters: 44

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	666	85.9	1251	12	US-10-399-645-7
2	666	85.9	1316	15	US-10-371-905A-16
3	661	85.3	762	14	US-10-163-547-1
4	112	14.5	112	9	US-09-864-761-47587
5	80	10.3	1326	15	US-10-379-981-12
6	66	8.5	66	9	US-09-864-761-46742
7	16	2.1	151	12	US-10-371-905A-48
8	16	2.1	248	12	US-10-371-905A-46
9	16	2.1	271	12	US-10-371-905A-44
10	16	2.1	314	15	US-10-274-639-11
11	16	2.1	314	16	US-10-333-574-11
12	16	2.1	357	12	US-10-371-905A-40
13	16	2.1	398	12	US-10-371-905A-42
14	16	2.1	472	16	US-10-408-765A-2752
15	16	2.1	530	10	US-09-372-348-11

Sequence 12, Appl
Sequence 16, Appl
Sequence 18, Appl
Sequence 30, Appl
Sequence 34, Appl
Sequence 2, Appl
Sequence 20, Appl
Sequence 14, Appl
Sequence 8, Appl
Sequence 28, Appl
Sequence 13, Appl
Sequence 15, Appl
Sequence 1, Appl
Sequence 9, Appl
Sequence 6, Appl
Sequence 26, Appl
Sequence 32, Appl
Sequence 36, Appl
Sequence 4, Appl
Sequence 243156,
Sequence 10, Appl
Sequence 276474,
Sequence 22, Appl
Sequence 132523,
Sequence 246594,
Sequence 38, Appl
Sequence 223768,
Sequence 47268, A

16 2.1 530 12 US-10-371-905A-12
17 2.1 530 12 US-10-371-905A-16
18 2.1 530 12 US-10-371-905A-18
19 2.1 530 12 US-10-371-905A-30
20 2.1 530 12 US-10-371-905A-34
21 2.1 530 14 US-10-107-695B-2
22 2.1 565 12 US-10-371-905A-20
23 2.1 574 12 US-10-371-905A-14
24 2.1 633 12 US-10-371-905A-24
25 2.1 1021 12 US-10-371-905A-8
26 1.9 497 12 US-10-371-905A-28
27 1.8 245 10 US-09-372-348-13
28 1.8 748 10 US-09-372-348-15
29 1.8 1123 12 US-09-972-525-1
30 1.7 360 10 US-09-372-348-9
31 1.7 530 12 US-10-371-905A-6
32 1.7 530 12 US-10-371-905A-26
33 1.7 530 12 US-10-371-905A-32
34 1.7 530 12 US-10-371-905A-36
35 1.7 1016 12 US-10-371-905A-4
36 1.5 150 12 US-10-424-599-243156
37 1.4 530 12 US-10-371-905A-10
38 1.4 623 12 US-10-424-599-276474
39 1.4 715 12 US-10-371-905A-22
40 1.4 844 16 US-10-437-963-132523
41 1.3 240 12 US-10-424-599-246594
42 1.3 417 12 US-10-371-905A-38
43 1.3 549 12 US-10-424-599-223768
44 1.3 1003 12 US-10-425-114-47268

ALIGNMENTS

RESULT 1

US-10-399-645-7
; Sequence 7, Application US/10399645
; Publication No. US20040029249A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; LEE, Ernestine A.
; APPLICANT: HAFALIA, April J.A.; YUE, Henry
; APPLICANT: LAL, Preeti G.; YAO, Monique G.
; APPLICANT: LU, Yan; CHAWLA, Narinder K.
; APPLICANT: WARREN, Bridget A.; LU, Dying Aina M.
; APPLICANT: BAUGHN, Mariah R.; DELEGARNE, Angelo M.
; APPLICANT: BUREFORD, Neil; BOROWSKY, Mark L.
; APPLICANT: LEE, Sally; XU, Yuming
; APPLICANT: GRIFFIN, Jennifer A.; KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Falda; ELLIOTT, Vicki S.
; APPLICANT: SWARNAKAR, Anita; RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Daniel B.; TRIBOULEY, Catherine M.
; APPLICANT: LO, Terence P.; AU-YOUNG, Janice K.
; APPLICANT: THANGAVELU, Kavitha; KEARNEY, Liam
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0263 USN
; CURRENT APPLICATION NUMBER: US/10/399,645
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: PCT/US01/51034
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 60/241,573
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/243,643
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/245,256
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/248,395
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/249,826
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,303
; PRIOR FILING DATE: 2000-11-20

; PRIORITY APPLICATION NUMBER: US 60/250,981
; PRIORITY FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 1251
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023249A1 7997881CD1
US-10-399-645-7

Query Match 85.9%; Score 666; DB 12; Length 1251;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 766; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSSEAVSPGMDAGSAGVSSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPSAYONQPGSSEAVSPGMDAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQKQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQKQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMMCTMWAHITQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMMCTMWAHITQALSNPGDVVKPMFVI 180
QY 181 NEMRRIARHLRFNGQEDAEHFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
DB 181 NEMRRIARHLRFNGQEDAEHFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSTDFDYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
DB 241 RVKCLNCKGVSTDFDYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
QY 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVIVLYAVL 360
DB 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVIVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLFYIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISORVVTNKQAAFGTGPQLPSHMIKNPPLHNGTGPKLDTFSSS 480
DB 421 LTHPTSPGQSSPRPVISORVVTNKQAAFGTGPQLPSHMIKNPPLHNGTGPKLDTFSSS 480
QY 481 MSSPNGSNVNRASPVNASAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 540
DB 481 MSSPNGSNVNRASPVNASAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 540
QY 541 LHSNSLENTPKVPVSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 600
DB 541 LHSNSLENTPKVPVSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 600
QY 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITIVSSHSPGQDAEDEATPHELQPMTLNGA 660
DB 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITIVSSHSPGQDAEDEATPHELQPMTLNGA 660
QY 661 NSADSDPKENGLAPDGASCQOPALHSENPPAKANGLPGLMPAPLLSLPDKKILETF 720
DB 661 NSADSDPKENGLAPDGASCQOPALHSENPPAKANGLPGLMPAPLLSLPDKKILETF 720
QY 721 RLNSKLGSTDEMSAPGAERGGPDRDAEPQPGSPAESLEBPDAAA 767
DB 721 RLNSKLGSTDEMSAPGAERGGPDRDAEPQPGSPAESLEBPDAAA 767

RESULT 2
US-10-379-981-9
; Sequence 9, Application US/10379981

; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE OF INVENTION: MURINE ORTHOLOG
; CURRENT APPLICATION NUMBER: US/10/379,981
; PRIORITY FILING DATE: 2003-03-05
; PRIORITY FILING DATE: 2002-09-08
; PRIORITY FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 9
; LENGTH: 1316
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-379-981-9

Query Match 85.9%; Score 666; DB 15; Length 1316;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 766; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSSEAVSPGMDAGSAGVSSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASSSDPSAYONQPGSSEAVSPGMDAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPKQKQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VVYSSSVDPKSPKQKQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMMCTMWAHITQALSNPGDVVKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHEHSKTCCHAEFCMMCTMWAHITQALSNPGDVVKPMFVI 180
QY 181 NEMRRIARHLRFNGQEDAEHFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
DB 181 NEMRRIARHLRFNGQEDAEHFLQYTVDMQKACLNGSNKLDHRTQATTLVCOIFGGYLR 240
QY 241 RVKCLNCKGVSTDFDYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
DB 241 RVKCLNCKGVSTDFDYLDITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSCCKKWVPA 300
QY 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVIVLYAVL 360
DB 301 SKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSOPNGEPIVIVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLFYIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISORVVTNKQAAFGTGPQLPSHMIKNPPLHNGTGPKLDTFSSS 480
DB 421 LTHPTSPGQSSPRPVISORVVTNKQAAFGTGPQLPSHMIKNPPLHNGTGPKLDTFSSS 480
QY 481 MSSPNGSNVNRASPVNASAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 540
DB 481 MSSPNGSNVNRASPVNASAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 540
QY 541 LHSNSLENTPKVPVSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 600
DB 541 LHSNSLENTPKVPVSTIITNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVVMNGSKLN 600
QY 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITIVSSHSPGQDAEDEATPHELQPMTLNGA 660
DB 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITIVSSHSPGQDAEDEATPHELQPMTLNGA 660
QY 661 NSADSDPKENGLAPDGASCQOPALHSENPPAKANGLPGLMPAPLLSLPDKKILETF 720
DB 661 NSADSDPKENGLAPDGASCQOPALHSENPPAKANGLPGLMPAPLLSLPDKKILETF 720

QY 721 RLSNKLKSGTDEMSAPGAERPPEDRDAEPQGSPPAAESLEEDPAAA 767
Db 721 RLSNKLKSGTDEMSAPGAERPPEDRDAEPQGSPPAAESLEEDPAAA 767

RESULT 3
US-10-163-547-1
; Sequence 1, Application US/10163547
; Publication No. US20030037350A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Myers, Rachel
; APPLICANT: Rudolf-Owen, Laura A.
; TITLE OF INVENTION: No. US20030037350A1el nucleic acid sequences encoding a
; TITLE OF INVENTION: human ubiquitin protease, lipase, dynamin, short chain
; TITLE OF INVENTION: dehydrogenase, and ADAM-TS Metalloprotease and uses
; TITLE OF INVENTION: therefore
; FILE REFERENCE: 35800/247836(5900-243)
; CURRENT APPLICATION NUMBER: US/10/163,547
; CURRENT FILING DATE: 2002-06-05
; PRIOR APPLICATION NUMBER: 09/407,356
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/704,918
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 09/435,311
; PRIOR FILING DATE: 1999-11-05
; PRIOR APPLICATION NUMBER: 09/796,100
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/185,503
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: 09/781,598
; PRIOR FILING DATE: 2001-02-12
; PRIOR APPLICATION NUMBER: 60/182,009
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: 09/782,952
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,408
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 09/496,005
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-163-547-1
Query Match 85.3%; Score 661; DB 14; Length 762;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MTIVDKASESDPSAYQNPQGSSEAVSGDMDAGSAGVSSINDVSNHTLSLGPVGA 60
Db 1 MTIVDKASESDPSAYQNPQGSSEAVSGDMDAGSAGVSSINDVSNHTLSLGPVGA 60
QY 61 VVYSSSVDPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 61 VVYSSSVDPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTPPLANYMLSHESKTCCHAEFCMWCMTQAHITQALSNPGDVIKPMFVI 180
Db 121 FANAALQCLTYTPPLANYMLSHESKTCCHAEFCMWCMTQAHITQALSNPGDVIKPMFVI 180
QY 181 NEMRRIARHLRGNEQDAHEFLQYTVDMQKACINGSNKLDRHTQATTLVCOIFGGYLR 240
Db 181 NEMRRIARHLRGNEQDAHEFLQYTVDMQKACINGSNKLDRHTQATTLVCOIFGGYLR 240
QY 241 RVKLCNCKGVSTDPYLDITLIEIKAAQSVNKALQFVKPEOLDGENSYKCSKCKMWP 300
Db 241 RVKLCNCKGVSTDPYLDITLIEIKAAQSVNKALQFVKPEOLDGENSYKCSKCKMWP 300

QY 301 SKRFTTHRSSNVLTSLKRFANFTGGKIAKVXPEYLDIRPYMSQNGRPIVVVLAVL 360
Db 301 SKRFTTHRSSNVLTSLKRFANFTGGKIAKVXPEYLDIRPYMSQNGRPIVVVLAVL 360
QY 361 VHTGFENCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQQAYVLFYIRSHDVKNNGE 420
Db 361 VHTGFENCHAGHYFCYIKASNGLWYQMNDSIVSTDIRSVLSQQAYVLFYIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQRVVTNKQAAPGFIGQLPSHMIKNPPHNLNGTGPLKDTFSSS 480
Db 421 LTHPTSPGQSSPRPVISQRVVTNKQAAPGFIGQLPSHMIKNPPHNLNGTGPLKDTFSSS 480
QY 481 MSSPENGSSVNRASPVNASASVQNVNSVNRSSVPEHPKKOKITISIHKLKPVRCQSQPN 540
Db 481 MSSPENGSSVNRASPVNASASVQNVNSVNRSSVPEHPKKOKITISIHKLKPVRCQSQPN 540
QY 541 LHSNLSLENPTKVPSSITTNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVMMGKSKLN 600
Db 541 LHSNLSLENPTKVPSSITTNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVMMGKSKLN 600
QY 601 SSVLPVYGAESSEDSDEESKGLKENGIGITIVSSHSPGQDAEDEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSEDSDEESKGLKENGIGITIVSSHSPGQDAEDEATPHELQEPMTLNGA 660
QY 661 NSADSDSDPKENGLAPDQASCOGAPALHSENPPAKANGLPGLKMPAPLLSLPEDKILETF 720
Db 661 NSADSDSDPKENGLAPDQASCOGAPALHSENPPAKANGLPGLKMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKSGTDEMSAPGAERPPEDRDAEPQGSPPAAESLEE 762
Db 721 RLSNKLKSGTDEMSAPGAERPPEDRDAEPQGSPPAAESLEE 762

RESULT 4
US-09-864-761-47587
; Sequence 47587, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

```

Db      190 RFGNQEDAHEFLQYTVDMQKACLNGSNKLDRHTQAATLVQCIFGGYLRSVKCLNCKGV 249

Qy      251 SDTDPDYLDITLEIKAAQSV 270
        |||||
Db      250 SDTDPDYLDITLEIKAAQSV 269

RESULT 6
US-09-864-761-46742
; Sequence 46742, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: JS 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: JS 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: JS 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46742
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.77

Query Match          14.5%; Score 112; DB 9; Length 112;
Best Local Similarity 100.0%; Pred. No. 2.3e-100;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      591 PVMNGSKLNSSVLVPYGAESSEDSDESGKLGKENGIGTIVSSHSPGQDAEDBEATPHE 650
        |||||
Db      1   PVMNGSKLNSSVLVPYGAESSEDSDESGKLGKENGIGTIVSSHSPGQDAEDBEATPHE 60

Qy      651 LQPMTINGANSADSDPKENGLAPDGASCQCPALHSENPFKANGILPGK 702
        |||||
Db      61 LQPMTINGANSADSDPKENGLAPDGASCQCPALHSENPFKANGILPGK 112

RESULT 5
US-10-379-981-12
; Sequence 12, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; PRIOR FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 1326
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-379-981-12

Query Match          10.3%; Score 80; DB 15; Length 1326;
Best Local Similarity 100.0%; Pred. No. 3.1e-68;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      191 RFGNQEDAHEFLQYTVDMQKACLNGSNKLDRHTQAATLVQCIFGGYLRSVKCLNCKGV 250

```

; OTHER INFORMATION: SWISSPROT HIT: Q61068, EVALUE 6.00e-14
; OTHER INFORMATION: EST_HUMAN HIT: AW504799.1, EVALUE 9.00e-35
US-09-864-761-46742

Query Match 8.5%; Score 66; DB 9; Length 66;
Best Local Similarity 100.0%; Pred. No. 8.3e-56;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKQQQTHRVAGLQNLGNTCFANAALQCLTYTPPLANYMLS 141
DB 1 LGDGIAPPQKVLFPSEKICLKQQQTHRVAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60

QY 142 HEHST 147
DB 61 HEHST 66

RESULT 7
US-10-371-905A-48
; Sequence 48, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 48
; LENGTH: 151
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.10
US-10-371-905A-48

Query Match 2.1%; Score 16; DB 12; Length 151;
Best Local Similarity 100.0%; Pred. No. 8e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
DB 95 LQCLTYTPPLANYMLS 110

RESULT 8
US-10-371-905A-46
; Sequence 46, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.9
US-10-371-905A-46

Query Match 2.1%; Score 16; DB 12; Length 248;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
DB 95 LQCLTYTPPLANYMLS 110

RESULT 9
US-10-371-905A-44
; Sequence 44, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.2
US-10-371-905A-44

Query Match 2.1%; Score 16; DB 12; Length 271;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
DB 129 LQCLTYTPPLANYMLS 144

RESULT 10
US-10-274-639-11
; Sequence 11, Application US/10274639
; Publication No. US20030232349A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DELEGANE, Angelo M.; GANDHI, Ameena R.
; APPLICANT: HAFALIA, April J.A.; LU, Dying Aina M.
; APPLICANT: PATTERSON, Chandra; TRIBOULEY, Catherine M.
; APPLICANT: DAS, Debopriya; KALLICK, Deborah A.
; APPLICANT: NGUYEN, Daniel B.; LEE, Ernestine A.
; APPLICANT: KHAN, Farrah A.; YUE, Henry
; APPLICANT: AU-YOUNG, Janice K.; GRIFFIN, Jennifer A.
; APPLICANT: POLICKY, Jennifer L.; RAMKOMAR, Jayalaxmi
; APPLICANT: YANG, Junning; THANGAVELU, Kavitha
; APPLICANT: DING, Li; KEARNEY, Liam
; APPLICANT: BAUGHN, Mariah R.; BOROWSKY, Mark L.
; APPLICANT: SANJANWALA, Madhusudan M.; YAO, Monique G.
; APPLICANT: BURFORD, Neil; WALIA, Narinder K.
; APPLICANT: LAL, Preeti G.; LEE, Sally
; APPLICANT: TODD, Stephen; LO, Terence P.
; APPLICANT: TANG, Y. Tom; ELLIOTT, Vicki S.
; APPLICANT: AZIMZAI, Yalda; LU, Yan
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0167 USA
; CURRENT APPLICATION NUMBER: US/10/274,639
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: PCT/US01/22397
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/220,063
; PRIOR FILING DATE: 2000-07-21
; PRIOR APPLICATION NUMBER: US 60/221,680
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,544
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/224,717

Qy 126 LQCLTYTPPLANYMLS 141
db 95 LQCLTYTPPLANYMLS 110

Query Match	Best Local Similarity	Score	DB	Length	Mismatches	Indels	Gaps
<p>RESULT 14</p> <p>US-10-408-765A-2752</p> <p>; Sequence 2752, Application US/10408765A</p> <p>; Publication No. US20040101874A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Ghosh, Soumitra S.</p> <p>; APPLICANT: Fahy, Eoin D.</p> <p>; APPLICANT: Zhang, Bing</p> <p>; APPLICANT: Gibson, Bradford W.</p> <p>; APPLICANT: Taylor, Steven W.</p> <p>; APPLICANT: Glenn, Gary M.</p> <p>; APPLICANT: Warnock, Dale E.</p> <p>; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION</p> <p>; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME</p> <p>; FILE REFERENCE: 66088.465</p> <p>; CURRENT APPLICATION NUMBER: US/10/408,765A</p> <p>; CURRENT FILING DATE: 2003-04-04</p> <p>; NUMBER OF SEQ ID NOS: 3077</p> <p>; SOFTWARE: FastSeq for Windows Version 4.0</p> <p>; SEQ ID NO 2752</p> <p>; LENGTH: 472</p> <p>; TYPE: PRT</p> <p>; ORGANISM: Homo sapiens</p> <p>US-10-408-765A-2752</p>							
Query Match	Best Local Similarity	2.1%;	Score 16;	DB 16;	Length 472;		
Matches	16;	Conservative	0;	Mismatches	0;	Indels	0;
QY	126	LOCLTYTPPLANM	141				
DB	95	LOCLTYTPPLANM	110				
<p>RESULT 15</p> <p>US-09-372-348-11</p> <p>; Sequence 11, Application US/09372348A</p> <p>; Publication No. US20030028005A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Hedrick, Joseph A.</p> <p>; APPLICANT: Wang, Luquan</p> <p>; APPLICANT: Zlotnik, Albert</p> <p>; APPLICANT: Murgolo, Nicholas J.</p> <p>; APPLICANT: Greene, Jonathan R.</p> <p>; APPLICANT: Johnston, James A.</p> <p>; APPLICANT: Bazan, J. Fernando</p> <p>; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods</p> <p>; FILE REFERENCE: EX084K(2d)</p> <p>; CURRENT APPLICATION NUMBER: US/09/372,348A</p> <p>; CURRENT FILING DATE: 1999-08-11</p> <p>; NUMBER OF SEQ ID NOS: 29</p> <p>; SOFTWARE: PatentIn Ver. 2.0</p> <p>; SEQ ID NO 11</p> <p>; LENGTH: 530</p> <p>; TYPE: PRT</p> <p>; ORGANISM: primate</p> <p>US-09-372-348-11</p>							
Query Match	Best Local Similarity	2.1%;	Score 16;	DB 10;	Length 530;		
Matches	16;	Conservative	0;	Mismatches	0;	Indels	0;
QY	126	LOCLTYTPPLANM	141				
DB	95	LOCLTYTPPLANM	110				
<p>RESULT 16</p> <p>US-10-371-905A-12</p> <p>; Sequence 12, Application US/10371905A</p> <p>; Publication No. US20030224969A1</p>							
<p>RESULT 17</p> <p>US-10-371-905A-16</p> <p>; Sequence 16, Application US/10371905A</p> <p>; Publication No. US20030224969A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Hahn, Chang</p> <p>; APPLICANT: Liu, Hong</p> <p>; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes</p> <p>; FILE REFERENCE: USAV2001/0185</p> <p>; CURRENT APPLICATION NUMBER: US/10/371,905A</p> <p>; CURRENT FILING DATE: 2003-02-20</p> <p>; NUMBER OF SEQ ID NOS: 54</p> <p>; SOFTWARE: PatentIn version 3.1</p> <p>; SEQ ID NO 16</p> <p>; LENGTH: 530</p> <p>; TYPE: PRT</p> <p>; ORGANISM: Homo sapiens</p> <p>; NAME/KEY: MISC FEATURE</p> <p>; OTHER INFORMATION: human DUB4.6</p> <p>US-10-371-905A-16</p>							
Query Match	Best Local Similarity	2.1%;	Score 16;	DB 12;	Length 530;		
Matches	16;	Conservative	0;	Mismatches	0;	Indels	0;
QY	126	LOCLTYTPPLANM	141				
DB	95	LOCLTYTPPLANM	110				
<p>RESULT 18</p> <p>US-10-371-905A-18</p> <p>; Sequence 18, Application US/10371905A</p> <p>; Publication No. US20030224969A1</p> <p>; GENERAL INFORMATION:</p> <p>; APPLICANT: Hahn, Chang</p> <p>; APPLICANT: Liu, Hong</p> <p>; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes</p> <p>; FILE REFERENCE: USAV2001/0185</p> <p>; CURRENT APPLICATION NUMBER: US/10/371,905A</p> <p>; CURRENT FILING DATE: 2003-02-20</p> <p>; NUMBER OF SEQ ID NOS: 54</p> <p>; SOFTWARE: PatentIn version 3.1</p> <p>; SEQ ID NO 18</p> <p>; LENGTH: 530</p>							

```
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.7
US-10-371-905A-18

Query Match
Best Local Similarity 100.0%; DB 12; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANMYS 141
Db 95 LOCLTYTPPLANMYS 110
|||||

RESULT 19
US-10-371-905A-30
; Sequence 30, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.5
US-10-371-905A-30

Query Match
Best Local Similarity 100.0%; DB 12; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANMYS 141
Db 95 LOCLTYTPPLANMYS 110
|||||

RESULT 20
US-10-371-905A-34
; Sequence 34, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 34
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.7
US-10-371-905A-34

Query Match
Best Local Similarity 100.0%; DB 12; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANMYS 141
Db 95 LOCLTYTPPLANMYS 110
|||||

RESULT 21
US-10-107-695B-2
; Sequence 2, Application US/10107695B
; Publication No. US20030022201A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 68999, Human Ubiquitin
; TITLE OF INVENTION: Carboxyl-Terminal Hydrolase Family Member and Uses Therefor
; FILE REFERENCE: MPI01-076PIRCPIM
; CURRENT APPLICATION NUMBER: US/10/107,695B
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 60/279,184
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION:
US-10-107-695B-2

Query Match
Best Local Similarity 100.0%; DB 14; Length 530;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANMYS 141
Db 95 LOCLTYTPPLANMYS 110
|||||

RESULT 22
US-10-371-905A-20
; Sequence 20, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.8
US-10-371-905A-20

Query Match
Best Local Similarity 100.0%; DB 12; Length 565;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LOCLTYTPPLANMYS 141
Db 130 LOCLTYTPPLANMYS 145
|||||

RESULT 23
US-10-371-905A-14
; Sequence 14, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
```

```
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 574
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.5
US-10-371-905A-14

Query Match      2.1%; Score 16; DB 12; Length 574;
Best Local Similarity 100.0%; Pred. No. 2.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 LQCLTYTPPLANYMLS 141
      |||||||
Db      130 LQCLTYTPPLANYMLS 145

RESULT 24
US-10-371-905A-24
; Sequence 24, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 633
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.11
US-10-371-905A-24

Query Match      2.1%; Score 16; DB 12; Length 633;
Best Local Similarity 100.0%; Pred. No. 3e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 LQCLTYTPPLANYMLS 141
      |||||||
Db      198 LQCLTYTPPLANYMLS 213

RESULT 25
US-10-371-905A-8
; Sequence 8, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.2a
US-10-371-905A-8

Query Match      2.1%; Score 16; DB 12; Length 1021;
Best Local Similarity 100.0%; Pred. No. 4.7e-06;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      126 LQCLTYTPPLANYMLS 141
      |||||||
Db      11 LQCLTYTPPLANYMLS 26

RESULT 26
US-10-371-905A-28
; Sequence 28, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.3
US-10-371-905A-28

Query Match      1.9%; Score 15; DB 12; Length 497;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      127 LQCLTYTPPLANYMLS 141
      |||||||
Db      96 LQCLTYTPPLANYMLS 110

RESULT 27
US-09-372-348-13
; Sequence 13, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Juquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; APPLICANT: Bazar, J. Fernando
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 245
; TYPE: PRT
; ORGANISM: Primate
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.13
US-09-372-348-13

Query Match      1.8%; Score 14; DB 10; Length 245;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      310 SNVLTSLKRFANF 323
```

```
Db          96 SNVLTSLKRPANF 109
|||||
RESULT 28
US-09-372-348-15
; Sequence 15, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 360
; TYPE: PRT
; ORGANISM: primate
US-09-372-348-9
Query Match          1.7%; Score 13; DB 10; Length 360;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          398 SVLSQQAYVLFYI 410
          |||||
Db          191 SVLSQQAYVLFYI 203
          |||||

RESULT 31
US-10-371-905A-6
; Sequence 6, Application US/10371905A
; Publication No. US2003022469A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.1b
US-10-371-905A-6
Query Match          1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          398 SVLSQQAYVLFYI 410
          |||||
Db          361 SVLSQQAYVLFYI 373
          |||||

RESULT 32
US-10-371-905A-26
; Sequence 26, Application US/10371905A
; Publication No. US2003022469A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26
; LENGTH: 530
; TYPE: PRT

Db          96 SNVLTSLKRPANF 109
|||||
RESULT 28
US-09-372-348-15
; Sequence 15, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 748
; TYPE: PRT
; ORGANISM: primate
US-09-372-348-15
Query Match          1.8%; Score 14; DB 10; Length 748;
Best Local Similarity 100.0%; Pred. No. 0.00031;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          310 SNVLTSLKRPANF 323
          |||||
Db          102 SNVLTSLKRPANF 115
          |||||

RESULT 29
US-09-972-525-1
; Sequence 1, Application US/09972525
; Publication No. US20020086407A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/972,525
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/408,865
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-972-525-1
Query Match          1.8%; Score 14; DB 12; Length 1123;
Best Local Similarity 100.0%; Pred. No. 0.00045;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          310 SNVLTSLKRPANF 323
          |||||
Db          321 SNVLTSLKRPANF 334
          |||||

RESULT 30
US-09-372-348-9
; Sequence 9, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; OTHER INFORMATION: human DUB8.1
US-10-371-905A-26

Query Match 1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410
Db 361 SVLSQAYVLFYI 373
|||||

RESULT 33

US-10-371-905A-32
; Sequence 32, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; OTHER INFORMATION: human DUB8.6
US-10-371-905A-32

Query Match 1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410
Db 361 SVLSQAYVLFYI 373
|||||

RESULT 34

US-10-371-905A-36
; Sequence 36, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; OTHER INFORMATION: human DUB8.8
US-10-371-905A-36

Query Match 1.7%; Score 13; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410

Db 361 SVLSQAYVLFYI 373
|||||

RESULT 35

US-10-371-905A-4
; Sequence 4, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC_FEATURE
; OTHER INFORMATION: human DUB4.1a
US-10-371-905A-4

Query Match 1.7%; Score 13; DB 12; Length 1016;
Best Local Similarity 100.0%; Pred. No. 0.0038;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 398 SVLSQAYVLFYI 410
Db 847 SVLSQAYVLFYI 859
|||||

RESULT 36

US-10-424-599-243156
; Sequence 243156, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 243156
; LENGTH: 150
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_61599C.1.pap
US-10-424-599-243156

Query Match 1.5%; Score 12; DB 12; Length 150;
Best Local Similarity 100.0%; Pred. No. 0.0062;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLAN 137
Db 39 LQCLTYTPPLAN 50
|||||

RESULT 37

US-10-371-905A-10
; Sequence 10, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang

; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.2b
; US-10-371-905A-10

Query Match 1.4%; Score 11; DB 12; Length 530;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 400 LSQQAYVLFYI 410
Db 363 LSQQAYVLFYI 373

RESULT 38

US-10-424-599-276474
; Sequence 276474, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 276474
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: Clone ID: PAT_MRT3847_91676C.1.pap
; US-10-424-599-276474

Query Match 1.4%; Score 11; DB 12; Length 623;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AGLQNLGNTCF 121
Db 16 AGLQNLGNTCF 26

RESULT 39

US-10-371-905A-22
; Sequence 22, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 715
; TYPE: PRT

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB4.10
; US-10-371-905A-22

Query Match 1.4%; Score 11; DB 12; Length 715;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 400 LSQQAYVLFYI 410
Db 548 LSQQAYVLFYI 558

RESULT 40

US-10-437-963-132523
; Sequence 132523, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 132523
; LENGTH: 844
; TYPE: PRT
; ORGANISM: Oryza sativa
; NAME/KEY: unsure
; LOCATION: (1)..(844)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_34484C.1.pap
; US-10-437-963-132523

Query Match 1.4%; Score 11; DB 16; Length 844;
Best Local Similarity 100.0%; Pred. No. 0.29;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GAGLQNLGNTC 120
Db 99 GAGLQNLGNTC 109

RESULT 41

US-10-424-599-246594
; Sequence 246594, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 246594
; LENGTH: 240
; TYPE: PRT

ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(240)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MFT3847_64705C.1.pep
US-10-424-599-246594

Query Match 1.3%; Score 10; DB 12; Length 240;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 112 GLONLGNCTCF 121
| | | | | | | | | |
Db 129 GLONLGNCTCF 138

RESULT 42

US-10-371-905A-38
; Sequence 38, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 38
; LENGTH: 417
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: human DUB8.11
US-10-371-905A-38

Query Match 1.3%; Score 10; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 353 VYLYAVLVH 362
| | | | | | | | | |
Db 224 VYLYAVLVH 233

RESULT 43

US-10-424-599-223768
; Sequence 223768, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223768
; LENGTH: 549
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: Clone ID: PAT_MFT3847_44092C.1.pep
US-10-424-599-223768

Query Match 1.3%; Score 10; DB 12; Length 549;

Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 193 GNQEDAHEFL 202
| | | | | | | | | |
Db 110 GNQEDAHEFL 119

RESULT 44

US-10-425-114-47268
; Sequence 47268, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 47268
; LENGTH: 1003
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: MISC FEATURE
; OTHER INFORMATION: Clone ID: 700856585_FLI.pep
US-10-425-114-47268

Query Match 1.3%; Score 10; DB 12; Length 1003;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 GNQEDAHEFL 202
| | | | | | | | | |
Db 565 GNQEDAHEFL 574

Search completed: August 10, 2004, 20:22:35
Job time : 52 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 20:14:12 ; Search time 21 Seconds
(without alignments)
1905.244 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 775
Sequence: 1 MTIVDKASSESDPSAYQNQP.....AARSLEPDAASLFPFSEG 775

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 10

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued Patents AA.*
1: /cgm2_6/ptodata/2/iaa/5A COMB.pap.*
2: /cgm2_6/ptodata/2/iaa/5B COMB.pap.*
3: /cgm2_6/ptodata/2/iaa/6A COMB.pap.*
4: /cgm2_6/ptodata/2/iaa/6B COMB.pap.*
5: /cgm2_6/ptodata/2/iaa/PCTUS COMB.pap.*
6: /cgm2_6/ptodata/2/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	14	1.8	1123	4	US-09-408-865-1
2	13	1.7	494	3	US-09-019-095A-24

ALIGNMENTS

RESULT 1
US-09-408-865-1
; Sequence 1, Application US/09408865A
; Patent No. 6329171
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/408,865A
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-408-865-1

Query Match 1.8%; Score 14; DB 4; Length 1123;
Best Local Similarity 100.0%; Pred. No. 0.00012;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 310 SNVLTSLKRFANF 323
| | | | | | | | | |
Db 321 SNVLTSLKRFANF 334

RESULT 2
US-09-019-095A-24
; Sequence 24, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-019-095A-24

Query Match 1.7%; Score 13; DB 3; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 129 LTYTPPLANYMLS 141
| | | | | | | | | |
Db 63 LTYTPPLANYMLS 75

Search completed: August 10, 2004, 20:17:40
Job time : 21 secs

THIS PAGE BLANK (USPTO)

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	2371.6	98.3	4659	13	US-10-359-645-22	Sequence 22, Appl
2	2339.2	97.0	2347	15	US-10-163-547-2	Sequence 2, Appl
3	2310	95.8	4483	13	US-10-302-172-390	Sequence 390, Appl
4	2299.6	95.3	3951	16	US-10-379-981-8	Sequence 8, Appl
5	1487.6	61.7	3981	16	US-10-379-981-11	Sequence 11, Appl
6	443	18.4	1593	13	US-10-371-905A-33	Sequence 33, Appl
7	436	18.1	1698	13	US-10-371-905A-19	Sequence 19, Appl
8	434.6	18.0	1763	15	US-10-107-695B-1	Sequence 1, Appl
9	433.4	18.0	1593	13	US-10-371-905A-17	Sequence 17, Appl
10	431.8	17.9	1593	13	US-10-371-905A-11	Sequence 11, Appl
11	431.2	17.9	3051	13	US-10-371-905A-3	Sequence 3, Appl
12	430.2	17.8	1590	15	US-10-107-695B-3	Sequence 3, Appl
13	430.2	17.8	1593	13	US-10-371-905A-15	Sequence 15, Appl
14	430.2	17.8	1593	13	US-10-371-905A-25	Sequence 25, Appl

```
; PRIOR APPLICATION NUMBER: US 60/252,303
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,981
; PRIOR FILING DATE: 2000-12-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 4659
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040029249A1 7997881CB1
; US-10-399-645-22

Query Match      98.3%; Score 2371.6; DB 13; Length 4659;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2374; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGGCTGTGTCGGGGCGGGCGGGCGGGCGGGGGGATGGAGCGCGCGAGCGGGT 60
DB 83 GGGCTGTGTCGGGGCGGGCGGGCGGGCGGGGGATGGAGCGCGCGAGCGGGT 142
QY 61 CAGAGTTGAACAATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCCATCAGCCCTAT 120
DB 143 CAGAGTTGAACAATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCCATCAGCCCTAT 202
QY 121 CAGAATCAGCCTGGCAGCTCCGAGGCGAGTCTCAGCTGGAGACATGAGATGAGGTTCTGCC 180
DB 203 CAGAATCAGCCTGGCAGCTCCGAGGCGAGTCTCAGCTGGAGACATGAGATGAGGTTCTGCC 262
QY 181 AGCTGGGGTCTGTGCTTCAATGAATGATGTGTCAAATCACACATTTCTTTAGGACCA 240
DB 263 AGCTGGGGTCTGTGCTTCAATGAATGATGTGTCAAATCACACATTTCTTTAGGACCA 322
QY 241 GTACCTGGTGTGTAGTTTATTCAGAGTTCATCTGACTGATGAATCAAAACCATCACCA 300
DB 323 GTACCTGGTGTGTAGTTTATTCAGAGTTCATCTGACTGATGAATCAAAACCATCACCA 382
QY 301 CAAAAGGATCAAGCCCTAGGTGATGGATCGCTCTCCACAGAAAGTCTTTTCCCATCT 360
DB 383 CAAAAGGATCAAGCCCTAGGTGATGGATCGCTCTCCACAGAAAGTCTTTTCCCATCT 442
QY 361 GAGAAGATTGTCTTAAGTGGCAACAACATCATAGAGTTGGAGCTGGGCTCCAGAAATTG 420
DB 443 GAGAAGATTGTCTTAAGTGGCAACAACATCATAGAGTTGGAGCTGGGCTCCAGAAATTG 502
QY 421 GGCAATACCTGTTTGTGCAATGACGACTGAGTGTGTTAACTACACACACCTCTTGCC 480
DB 503 GGCAATACCTGTTTGTGCAATGACGACTGAGTGTGTTAACTACACACACCTCTTGCC 562
QY 481 AATTACATGCTATCACATGAACACTCCAAAACATGTCATGAGAGGCTTTTGTATGATG 540
DB 563 AATTACATGCTATCACATGAACACTCCAAAACATGTCATGAGAGGCTTTTGTATGATG 622
QY 541 TGTACAAATGCAACACATATTACCCAGGCTCATGTAATCTGGGGAGGTTATTAACCA 600
DB 623 TGTACAAATGCAACACATATTACCCAGGCTCATGTAATCTGGGGAGGTTATTAACCA 682
QY 601 ATGTTTGTCAATGAGATCGGGCTATAGCTAGGCACTCTCCGTTTGTGAAACCAAGAA 660
DB 683 ATGTTTGTCAATGAGATCGGGCTATAGCTAGGCACTCTCCGTTTGTGAAACCAAGAA 742
QY 661 GATGCCCATGAATTCCTTCAATACATCTGTTGATGTGAGAAAGCATGCTGAATGGC 720
DB 743 GATGCCCATGAATTCCTTCAATACATCTGTTGATGTGAGAAAGCATGCTGAATGGC 802
QY 721 AGCAATAAATTAGACAGACACACCCAGCCACCACTCTGTTGTCAGATATTTGGAGGA 780
DB 803 AGCAATAAATTAGACAGACACACCCAGCCACCACTCTGTTGTCAGATATTTGGAGGA 862
QY 781 TACCTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGCGGCTTCAGATACCTTTGATCCA 840
```

```
DB 863 TACCTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGCGGCTTTTCAGATACCTTTTATGCCA 922
QY 841 TATCTTTGATATAACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG 900
DB 923 TATCTTTGATATAACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG 982
QY 901 TTTGTGAAGCGCGGAAACAGCTTGTATGGAGAAAACCTCGTACAAAGTGCAGCAAGTGTAAAAAG 960
DB 983 TTTGTGAAGCGCGGAAACAGCTTGTATGGAGAAAACCTCGTACAAAGTGCAGCAAGTGTAAAAAG 1042
QY 961 ATGTTTCCAGCTTCAAAAGAGTTTCATATCCATAGATTCCTCTAATGTTCTTACATTTCT 1020
DB 1043 ATGTTTCCAGCTTCAAAAGAGTTTCATATCCATAGATTCCTCTAATGTTCTTACATTTCT 1102
QY 1021 CTGAACAGTTTTCGAAAATTTTACCGGTGGAAAATTTGCTAAGGATGTGAATATCCCTCTGAG 1080
DB 1103 CTGAACAGTTTTCGAAAATTTTACCGGTGGAAAATTTGCTAAGGATGTGAATATCCCTCTGAG 1162
QY 1081 TATCTTTGATATTCGGGCCATATATGTCTCAACCCAAACGAGAGCAATTTGTCTAGCTCTTG 1140
DB 1163 TATCTTTGATATTCGGGCCATATATGTCTCAACCCAAACGAGAGCAATTTGTCTAGCTCTTG 1222
QY 1141 TATGCAAGTGTGTCACACTGGTTTTTAATGCGATGCTGCGCATTTACTTCTGCTACATA 1200
DB 1223 TATGCAAGTGTGTCACACTGGTTTTTAATGCGATGCTGCGCATTTACTTCTGCTACATA 1282
QY 1201 AAAGCTAGCAATGGCTCTGCTGATCAAAATGAATGACTCCCATTTGTATCTACAGTGATATT 1260
DB 1283 AAAGCTAGCAATGGCTCTGCTGATCAAAATGAATGACTCCCATTTGTATCTACAGTGATATT 1342
QY 1261 AGATCGGTAATCTAGCCAAACAGCTATGTCTCTTTTATATCAGGTCCTCATGATGTGAAA 1320
DB 1343 AGATCGGTAATCTAGCCAAACAGCTATGTCTCTTTTATATCAGGTCCTCATGATGTGAAA 1402
QY 1321 AATGGAGGTGAACTTACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTC 1380
DB 1403 AATGGAGGTGAACTTACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGCTC 1462
QY 1381 ATCAGTCAGCGGGTGTGCACCAACACAGCTGCGCCAGGCTTTATCGGACCAACAGCTT 1440
DB 1463 ATCAGTCAGCGGGTGTGCACCAACACAGCTGCGCCAGGCTTTATCGGACCAACAGCTT 1522
QY 1441 CCCTCTCACATGATTAAGAATCCACCTCACTTAAATGGGAGTGGACCAATTTGAAAGACACG 1500
DB 1523 CCCTCTCACATGATTAAGAATCCACCTCACTTAAATGGGAGTGGACCAATTTGAAAGACACG 1582
QY 1501 CCAAGCAGTTCCTATGTCAGTCTTAAACGGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTT 1560
DB 1583 CCAAGCAGTTCCTATGTCAGTCTTAAACGGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTT 1642
QY 1561 AATGCTTTCAGCTTCTGTCGCAAACTGTCAGTTAATAGGTCCTCAGTGTATCCAGAACAT 1620
DB 1643 AATGCTTTCAGCTTCTGTCGCAAACTGTCAGTTAATAGGTCCTCAGTGTATCCAGAACAT 1702
QY 1621 CCTAAGAAAACAAAATTTACAATCAGTATTCAACAACAAAGTTGCTGTTGCGCAGGTGAC 1680
DB 1703 CCTAAGAAAACAAAATTTACAATCAGTATTCAACAACAAAGTTGCTGTTGCGCAGGTGAC 1762
QY 1681 TCTCAACCTTAACCTTCAATGATTAATCTTTGGAGAAACCTTCAACAGCCGCTTCCCTCTTCT 1740
DB 1763 TCTCAACCTTAACCTTCAATGATTAATCTTTGGAGAAACCTTCAACAGCCGCTTCCCTCTTCT 1822
QY 1741 ACCAATTAACCAATTCGAGTACAGTCTACCTCGAAGCATCTACGATGTCTAGTCTTCTAGT 1800
DB 1823 ACCAATTAACCAATTCGAGTACAGTCTACCTCGAAGCATCTACGATGTCTAGTCTTCTAGT 1882
QY 1801 AAAGTAACAAAACCAATCCCCCGCAGTGAATCTGTGCTCCAGCCGCTGATGAATGGCAAA 1860
DB 1883 AAAGTAACAAAACCAATCCCCCGCAGTGAATCTGTGCTCCAGCCGCTGATGAATGGCAAA 1942
QY 1861 TCCAAAGCTGACTCAGCGTGTGCTGCTTATGGCCGAGTCTCTCTGAGACTCTGAC 1920
DB 1943 TCCAAAGCTGACTCAGCGTGTGCTGCTTATGGCCGAGTCTCTCTGAGACTCTGAC 2002
```


Db 844 TTGGAGTAAGAGCTGCTCAGAGTGTCACAAAGCAATGGAGCAGTTTGTGAAGCCGGA 903
QY 916 CAGCTTTGATGGAGAAAACCTCGTCAAGTGCAGCAAGTGTAAAAGATGGTTCCAGCTTCA 975
Db 904 CAGCTTTGATGGAGAAAACCTCGTCAAGTGCAGCAAGTGTAAAAGATGGTTCCAGCTTCA 963
QY 976 AAGAGGTTCACTATCCATAGATCCTCTAATGTGTTCTTACACTTTCTCTGAACGTTTGGCA 1035
Db 964 AAGAGGTTCACTATCCATAGATCCTCTAATGTGTTCTTACACTTTCTCTGAACGTTTGGCA 1023
QY 1036 AATTTTACCGGTGGAAAAATTTGCTAAGGATGTGAATACCCCTCAGTATCTTTGATATTCGG 1095
Db 1024 AATTTTACCGGTGGAAAAATTTGCTAAGGATGTGAATACCCCTCAGTATCTTTGATATTCGG 1083
QY 1096 CCATATATGTTCTCAACCCAAACGAGAGCCAAATTTGTCTACGTTCTTTGATGCAAGTCTGGTC 1155
Db 1084 CCATATATGTTCTCAACCCAAACGAGAGCCAAATTTGTCTACGTTCTTTGATGCAAGTCTGGTC 1143
QY 1156 CACACTGGTTTAAATGGCCATGCTGGCCATTACTCTCCGCGCCCGCTCATCACTCAGCGGTT 1395
Db 1144 CACACTGGTTTAAATGGCCATGCTGGCCATTACTCTCCGCGCCCGCTCATCACTCAGCGGTT 1383
QY 1396 GTCCACCAACAAACAGGCTGCGCCAGGCTTTATCGGACCAAGCTTCCTCTCCATGATA 1455
Db 1384 GTCCACCAACAAACAGGCTGCGCCAGGCTTTATCGGACCAAGCTTCCTCTCCATGATA 1443
QY 1456 AAGAAATCCAGCTCACTTAAATGGAGCTGGACCAATTTGAAGACACGCGCAAGCTTCCATG 1515
Db 1444 AAGAAATCCAGCTCACTTAAATGGAGCTGGACCAATTTGAAGACACGCGCAAGCTTCCATG 1503
QY 1516 TCAGGTCCTAAACGGGAATTCAGGTGTCACAGAGCTAGTCTGTTAATGCTTCAGGTTCT 1575
Db 1504 TCAGGTCCTAAACGGGAATTCAGGTGTCACAGAGCTAGTCTGTTAATGCTTCAGGTTCT 1563
QY 1576 GTCCAAAACTGGTCAAGTTAATAGTCTCTCAGTGATCCAGAACATCTTAAGAAACAAAA 1635
Db 1564 GTCCAAAACTGGTCAAGTTAATAGTCTCTCAGTGATCCAGAACATCTTAAGAAACAAAA 1623
QY 1636 ATTACAAATCAGTATTCACAAAGATTGCTGTTGCGCCAGTGTCAAGCTCAACCTAACCTT 1695
Db 1624 ATTACAAATCAGTATTCACAAAGATTGCTGTTGCGCCAGTGTCAAGCTCAACCTAACCTT 1683
QY 1696 CATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTAACCAATTC 1755
Db 1684 CATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTAACCAATTC 1743
QY 1756 GCAGTACAGTCTACCTCGAAACGATCTACGATGTCAAGTTTCTAGTAAAGTAAACAAACCG 1815
Db 1744 GCAGTACAGTCTACCTCGAAACGATCTACGATGTCAAGTTTCTAGTAAAGTAAACAAACCG 1803
QY 1816 ATCCCCCGAGTGAATCTCTGCTCCAGCCGTTGATGAATGGCAAAATCCAGCTGAACCTC 1875
Db 1804 ATCCCCCGAGTGAATCTCTGCTCCAGCCGTTGATGAATGGCAAAATCCAGCTGAACCTC 1863
QY 1876 AGCGTGTGTCGCTATGCGCGCGAGTCTCTGAGGACTCTCAGCAGGAGTCAAAAGGGG 1935
Db 1864 AGCGTGTGTCGCTATGCGCGCGAGTCTCTGAGGACTCTCAGCAGGAGTCAAAAGGGG 1923
QY 1936 CTGGGAGAGAGATGGGATTTGTTACGATTTGAGCTCCACTCTCCGCGCCCAAGATGCC 1995

Db 1924 CTGGCAAGGAGAAATGGGATTGGTACGATTGTGAGCTCCCACTCTCCGCGCAAGATGCC 1983
QY 1996 GAAGATGAGAGAGGCCACTTCGCAAGAGCTTCAAGAAACCCATGACCTTAAACGGTGTCTAAT 2055
Db 1984 GAAGATGAGAGAGGCCACTTCGCAAGAGCTTCAAGAAACCCATGACCTTAAACGGTGTCTAAT 2043
QY 2056 AGTGACAGACGACACAGTGTGACCCGAAAGAAACGGCTAGCGCTGATGTGCGCAGCTGC 2115
Db 2044 AGTGACAGACGACACAGTGTGACCCGAAAGAAACGGCTAGCGCTGATGTGCGCAGCTGC 2103
QY 2116 CAAGGCGAGCTCCCTGCACCTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCTCTGGA 2175
Db 2104 CAAGGCGAGCTCCCTGCACCTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCTCTGGA 2163
QY 2176 AAGTTGATGCTCTCTCTTTGCTGTCTCTCCAGAGACAAATCTTTAGAGACCTTTCAGG 2235
Db 2164 AAGTTGATGCTCTCTCTTTGCTGTCTCTCCAGAGACAAATCTTTAGAGACCTTTCAGG 2223
QY 2236 CTTAGCAACAAACTGAAAGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGC 2295
Db 2224 CTTAGCAACAAACTGAAAGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGC 2283
QY 2296 CTTCCGAGACCGCGACGCGAGCTCAGCTGGAGCCCCCGCCGCAATCCCTGGAG 2355
Db 2284 CTTCCGAGACCGCGACGCGAGCTCAGCTGGAGCCCCCGCCGCAATCCCTGGAG 2343
QY 2356 GAGC 2359
Db 2344 GAGC 2347

RESULT 3
US-10-302-172-390
; Sequence 390, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20040053250A1 Arginine-rich Protein-like Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 950
; SOFTWARE: Pt_Fl_genes version 2.0
; SEQ ID NO 390
; LENGTH: 4483
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (191)..(4000)
US-10-302-172-390

Query Match 95.8%; Score 2310; DB 13; Length 4483;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2313; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 61 CAGAGTTGAACAATACCATAGTTGACAAAGCTTCTGAATCTTTCAGACCATCAGCCTAT 120
Db 179 CAGAGTTGAACAATACCATAGTTGACAAAGCTTCTGAATCTTTCAGACCATCAGCCTAT 238
QY 121 CAGAAATCAGCTGGCAGCTCCGAGCAGTCTCACCTGGAGACATGGATGAGGTTCTGCC 180
Db 239 CAGAAATCAGCTGGCAGCTCCGAGCAGTCTCACCTGGAGACATGGATGAGGTTCTGCC 298

QY 181 AGCTGGGTGCTGTGCTCTTCAATGATGTCGCAATCACAACATCTTCTTTAGGACCA 240
Db 299 AGCTGGGTGCTGTGCTCTTCAATGATGTCGCAATCACAACATCTTCTTTAGGACCA 358
QY 241 GTACTCTGTGCTGTAGTTTATTTACGAGTTCACTGTACCTGATATAAATCAAAACCATCACCA 300
Db 359 GTACTCTGTGCTGTAGTTTATTTACGAGTTCACTGTACCTGATATAAATCAAAACCATCACCA 418
QY 301 CAAAGGATCAAGCCCTAGGTGATGGATCGCTCTCTCAAGAAAGTTCTTTTCCCATCT 360
Db 419 CAAAGGATCAAGCCCTAGGTGATGGATCGCTCTCTCAAGAAAGTTCTTTTCCCATCT 478
QY 361 GAGAGATTCTCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTG 420
Db 479 GAGAGATTCTCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTG 538
QY 421 GGCAATACCTGTTTGGCAATGCAGCACTGCAGTGTTTAACTACACACCACTCTTGCC 480
Db 539 GGCAATACCTGTTTGGCAATGCAGCACTGCAGTGTTTAACTACACACCACTCTTGCC 598
QY 481 AATTACATGCTATCACATGAACACTCCAAACATGTCATGCGAAGGCTTTTGTATGATG 540
Db 599 AATTACATGCTATCACATGAACACTCCAAACATGTCATGCGAAGGCTTTTGTATGATG 658
QY 541 TGTACAATGCAAGCACATATTACCCAGGCACTCAGTAATCCTGGGGAGCTTATTAAACCA 600
Db 659 TGTACAATGCAAGCACATATTACCCAGGCACTCAGTAATCCTGGGGAGCTTATTAAACCA 718
QY 601 ATGTTTGTCAATAGATGCGGGTATGCTAGGCACTCCGTTTGGAAACCAAGAA 660
Db 719 ATGTTTGTCAATAGATGCGGGTATGCTAGGCACTCCGTTTGGAAACCAAGAA 778
QY 661 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTGATGGC 720
Db 779 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCTGAAAGCATGCTGATGGC 838
QY 721 AGCAATAATTAGACAGACACACCCAGGCCACCACTCTGTTTGTGCAGATATTCGAGGA 780
Db 839 AGCAATAATTAGACAGACACACCCAGGCCACCACTCTGTTTGTGCAGATATTCGAGGA 898
QY 781 TACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGCTTTCAGATCTTTGATCCA 840
Db 899 TACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGCTTTCAGATCTTTGATCCA 958
QY 841 TATCTTGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCACCAAGGCATTGGAGCG 900
Db 959 TATCTTGATATAACATTTGGAGATAAAGGCTGCTCAGAGTGTCACCAAGGCATTGGAGCG 1018
QY 901 TTTGTGAAGCCGGAACAGCTTGATGGAGAAAACCTCGTACAAGTGCGAGCAAGTGAAGAAG 960
Db 1019 TTTGTGAAGCCGGAACAGCTTGATGGAGAAAACCTCGTACAAGTGCGAGCAAGTGAAGAAG 1078
QY 961 ATGGTTCAGCTTCAAAGAGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTTCT 1020
Db 1079 ATGGTTCAGCTTCAAAGAGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTTCT 1138
QY 1021 CTGAACAGTTTTCGAAATTTACCGGTGAAATAATTGCTAAGAGATGTGAATACCCCTGAG 1080
Db 1139 CTGAACAGTTTTCGAAATTTACCGGTGAAATAATTGCTAAGAGATGTGAATACCCCTGAG 1198
QY 1081 TATCTTGATATTCGGCCATATATGCTCAACCCACCGAGAGCAATGCTACGTCCTTG 1140
Db 1199 TATCTTGATATTCGGCCATATATGCTCAACCCACCGAGAGCCAAATGCTACGTCCTTG 1258
QY 1141 TATGCACTGCTGGTCCACACTGGTTTAAATGCGATGCTGGCCATTACTTCTGCTACATA 1200
Db 1259 TATGCACTGCTGGTCCACACTGGTTTAAATGCGATGCTGGCCATTACTTCTGCTACATA 1318
QY 1201 AAAGCTAGCAATGGGCTCTGGTATCAAAATGAATGACTCCATTTGATACAGTGAATTT 1260
Db 1319 AAAGCTAGCAATGGGCTCTGGTATCAAAATGAATGACTCCATTTGATACAGTGAATTT 1378
QY 1261 AGATCGGTACTCAGCCAAACAGCCATGTCCTTTTATATCAGGTCCCATGATGTGAAA 1320

Db 1379 AGATCGGTACTCAGCCAAACAGCCATATGTGCTCTTTTATATCAGTCCCATGATGTAAA 1438
QY 1321 AATGGAGTGAACCTTACTATCCCAATGATGCTGCTCTCTCCCGCCGCTC 1380
Db 1439 AATGGAGTGAACCTTACTATCCCAATGATGCTGCTCTCTCCCGCCGCTC 1498
QY 1381 ATCAGTCAAGCGGTGTCACCAACAAACAGGTGCGCCAGGCTTTATCGGACCAAGCTT 1440
Db 1499 ATCAGTCAAGCGGTGTCACCAACAAACAGGTGCGCCAGGCTTTATCGGACCAAGCTT 1558
QY 1441 CCCTCTCATATGATAAAGAATCCACCTCACTTAAATGGGACTGGACCATTTGAAAGACACG 1500
Db 1559 CCCTCTCATATGATAAAGAATCCACCTCACTTAAATGGGACTGGACCATTTGAAAGACACG 1618
QY 1501 CCAAGCAGTTCATGTCAGTCTTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTT 1560
Db 1619 CCAAGCAGTTCATGTCAGTCTTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTT 1678
QY 1561 AATGCTTCAGCTTCTGTCGCAAACTGCTCAATTAATAGTCTCTCAGTATCCCAAGACAT 1620
Db 1679 AATGCTTCAGCTTCTGTCGCAAACTGCTCAATTAATAGTCTCTCAGTATCCCAAGACAT 1738
QY 1621 CCTAAGAAACAAAAATTTACAAATCAGTATTCACAAACAAGTTGCTCTGTCGCGAGTGTG 1680
Db 1739 CCTAAGAAACAAAAATTTACAAATCAGTATTCACAAACAAGTTGCTCTGTCGCGAGTGTG 1798
QY 1681 TCTCAACCTAACCTTCTAGTAAATTTCTTTGGAGAACCTCAACAGCCGCTTCCCTCTTCT 1740
Db 1799 TCTCAACCTAACCTTCTAGTAAATTTCTTTGGAGAACCTCAACAGCCGCTTCCCTCTTCT 1858
QY 1741 ACCATTACCAATTCGCAAGTACAGTCTACCTCGAAGCATCTACGATGTCAGTCTTCTAGT 1800
Db 1859 ACCATTACCAATTCGCAAGTACAGTCTACCTCGAAGCATCTACGATGTCAGTCTTCTAGT 1918
QY 1801 AAAGTAAACAAAACCATCCCGCAGTGAATCTCTCTCCAGCCGCTGATGAATGGCAAA 1860
Db 1919 AAAGTAAACAAAACCATCCCGCAGTGAATCTCTCTCCAGCCGCTGATGAATGGCAAA 1978
QY 1861 TCCAAAGTGAACCTCAGCGTGTGTCGCTTATGCGCGGAGTCTCTGAGAGTCTGAC 1920
Db 1979 TCCAAAGTGAACCTCAGCGTGTGTCGCTTATGCGCGGAGTCTCTCTGAGAGTCTGAC 2038
QY 1921 GAGGAGTCAAAAGGGCTCGGCAAGGAGAAATGGATTGTGAGTGTGAGTCTCCACTCT 1980
Db 2039 GAGGAGTCAAAAGGGCTCGGCAAGGAGAAATGGATTGTGAGTGTGAGTCTCCACTCT 2098
QY 1981 CCGCGCCCAAGATGCCGAAGATGAGGAGGCCACTCCGCGCAGAGCTTCAAGAACCCATGACC 2040
Db 2099 CCGCGCCCAAGATGCCGAAGATGAGGAGGCCACTCCGCGCAGAGCTTCAAGAACCCATGACC 2158
QY 2041 CTAAACCGGTGTATAATGATGCGAGACAGTACCGCAAGAAAGAAACGGCTTAGCGCT 2100
Db 2159 CTAAACCGGTGTATAATGATGCGAGACAGTACCGCAAGAAAGAAACGGCTTAGCGCT 2218
QY 2101 GATGCTGCGAGTCCCAAGGGCAGCTGCTCGCTGCACTCAGAAAAATCCCTTTCTTAAGGCA 2160
Db 2219 GATGCTGCGAGTCCCAAGGGCAGCTGCTCGCTGCACTCAGAAAAATCCCTTTCTTAAGGCA 2278
QY 2161 AACGCTCTTCTTGGAAAAGTTGATGCTGCTCTCTTGTGCTCTCTCCAGAGACAAAAATC 2220
Db 2279 AACGCTCTTCTTGGAAAAGTTGATGCTGCTCTCTCTTGTGCTCTCTCCAGAGACAAAAATC 2338
QY 2221 TTAGAGACCTTCAGGCTTAGCAACAAATGAAGGCTCGACGGATGAATGATGTCACCT 2280
Db 2339 TTAGAGACCTTCAGGCTTAGCAACAAATGAAGGCTCGACGGATGAATGATGTCACCT 2398
QY 2281 GGAGCAGAGAGGGGCTCCCGAGAGCCGAGCCGAGCTCAGCTGCGAGCCCGCC 2340
Db 2399 GGAGCAGAGAGGGGCTCCCGAGAGCCGAGCCGAGCTCAGCTGCGAGCCCGCC 2458
QY 2341 GCCGAATCCTCGAGAGCCAGATGCGGCGCAAGCTT 2378

Db 2459 GCCGAATCCCTGGAGGAGCCAGATGCGCGCGGCT 2496

RESULT 4

US-10-379-981-8
; Sequence 8, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE REFERENCE: USAV2002-0022, WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 3951
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-379-981-8

Query Match 95.3%; Score 2299.6; DB 16; Length 3951;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 2302; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	73	ATGACCATAGTTCACAAAGCTTCTGATCTTCAGACCCATCAGCCTATCAGATCAGCCT	132
Db	1	ATGACCATAGTTCACAAAGCTTCTGATCTTCAGACCCATCAGCCTATCAGATCAGCCT	60
QY	133	GGCAGCTCCGAGGAGCTCTCACCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGT	192
Db	61	GGCAGCTCCGAGGAGCTCTCACCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGT	120
QY	193	GTGCTTTCATGAATGATGTCAATTCACACTTCTTTTAGGACGAGTACCTGTGTGT	252
Db	121	GTGCTTTCATGAATGATGTCAATTCACACTTCTTTTAGGACGAGTACCTGTGTGT	180
QY	253	GTAGTTTATTCGAGTTCATCTGTACCTGTAAATCAAAACCATCACCAAAAGGATCAA	312
Db	181	GTAGTTTATTCGAGTTCATCTGTACCTGTAAATCAAAACCATCACCAAAAGGATCAA	240
QY	313	GCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCCTTTCCCATCTGAGAAAGTTGT	372
Db	241	GCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCCTTTCCCATCTGAGAAAGTTGT	300
QY	373	CTTAAGTGGCAACAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGCAATACCTGT	432
Db	301	CTTAAGTGGCAACAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGCAATACCTGT	360
QY	433	TTTGCCAAATGCAGCTGCAGTCTTTAACTCACACCACTCTTGCCAAATTCATGCTA	492
Db	361	TTTGCCAAATGCAGCTGCAGTCTTTAACTCACACCACTCTTGCCAAATTCATGCTA	420
QY	493	TCACATGAACACTCCAAAACATGTCATGCAAGAGGCTTTTGTATGATGTACAATGCCAA	552
Db	421	TCACATGAACACTCCAAAACATGTCATGCAAGAGGCTTTTGTATGATGTACAATGCCAA	480
QY	553	GCACATATTACCCAGCACTCAGTATCTGGGAGCTTTTAAACCAATGTTTGTATC	612
Db	481	GCACATATTACCCAGCACTCAGTATCTGGGAGCTTTTAAACCAATGTTTGTATC	540
QY	613	AATGAGATCGGGGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCATGAA	672
Db	541	AATGAGATCGGGGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCATGAA	600
QY	673	TTCCCTTCAATACACTGTTGATGCTATGCAAGAAAGCATGCTTGAATGGCAATAAATTA	732

Db	601	TTCTTCAATACACTGTTGATGCTATGCAAGAACATGCTTGAATGGCAGCAATAAATTA	660
QY	733	GACAGACACACCCAGCCACCACTCTTTGTTGTGACATATTTGGAGGATACCTAAGATCT	792
Db	661	GACAGACACACCCAGCCACCACTCTTTGTTGTGACATATTTGGAGGATACCTAAGATCT	720
QY	793	AGAGTCAATGTTTAAATTCGAAGGGGCTTTGAGATACTTTTGATCCATATCTTGATATA	852
Db	721	AGAGTCAATGTTTAAATTCGAAGGGGCTTTGAGATACTTTTGATCCATATCTTGATATA	780
QY	853	ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAACCG	912
Db	781	ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAACCG	840
QY	913	GAAACAGCTTCATGGAGAAAACTCGTACAAAGTCAGCAAGTGTAAAAAGATGGTTCAGCT	972
Db	841	GAAACAGCTTCATGGAGAAAACTCGTACAAAGTGTAAAAAGATGGTTCAGCT	900
QY	973	TCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCTCTGAAAACGTTT	1032
Db	901	TCAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCTCTGAAAACGTTT	960
QY	1033	GCAAAATTTACCGGTGGAAAAATTTGCTAAGGATGTGAAATACCTCGAGTATCTTGATAT	1092
Db	961	GCAAAATTTACCGGTGGAAAAATTTGCTAAGGATGTGAAATACCTCGAGTATCTTGATAT	1020
QY	1093	CGGCCATATATGCTCAACCCAAAGGAGCCAAATTTGCTACGTTCTTGTATGAGTGTCTG	1152
Db	1021	CGGCCATATATGCTCAACCCAAAGGAGCCAAATTTGCTACGTTCTTGTATGAGTGTCTG	1080
QY	1153	GTCCACACTGTTTAAATTTGCCATGCTGGCCATTACTTTCTGTACATAAAAGTAGCAAT	1212
Db	1081	GTCCACACTGTTTAAATTTGCCATGCTGGCCATTACTTTCTGTACATAAAAGTAGCAAT	1140
QY	1213	GGCTCTGGTATCAAAATGAATGATCTCAATGTATACCAAGTATATAGATCGGTACTC	1272
Db	1141	GGCTCTGGTATCAAAATGAATGATCTCAATGTATACCAAGTATATAGATCGGTACTC	1200
QY	1273	AGCCAAACAGCTTATGTGCTTTTATATCAGTCCCATGATGTGAAAAATGAGGTGAA	1332
Db	1201	AGCCAAACAGCTTATGTGCTTTTATATCAGTCCCATGATGTGAAAAATGAGGTGAA	1260
QY	1333	CTTACTCATCCACCCATAGCCCGCCGCTCTCTCCCGCCGCTCATCAGTCAGCGG	1392
Db	1261	CTTACTCATCCACCCATAGCCCGCCGCTCTCTCTCCCGCCGCTCATCAGTCAGCGG	1320
QY	1393	GTGTGTCAACAAACAAAGGCTGCGCCAGGCTTTTATCGGACCAACAGCTTCCCTCTCAGATG	1452
Db	1321	GTGTGTCAACAAACAAAGGCTGCGCCAGGCTTTTATCGGACCAACAGCTTCCCTCTCAGATG	1380
QY	1453	ATAAGAAATCCACCTCACTTAAATGGGACTGGACCAATTTGAAAGACACGCGCAAGCTTC	1512
Db	1381	ATAAGAAATCCACCTCACTTAAATGGGACTGGACCAATTTGAAAGACACGCGCAAGCTTC	1440
QY	1513	ATGTGAGTCCCTAACCGGAAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGCT	1572
Db	1441	ATGTGAGTCCCTAACCGGAAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGCT	1500
QY	1573	TCGTGCCAAATCGGTTCAGTTAATAGGTCCTCAGTGATCCAGAAACATCTTAAGAAACAA	1632
Db	1501	TCGTGCCAAATCGGTTCAGTTAATAGGTCCTCAGTGATCCAGAAACATCTTAAGAAACAA	1560
QY	1633	AAATTTACAAATCAGTATTCACAAAGTTGCTGTTGCGCAGTGTGAGTCTCAACCTAC	1692
Db	1561	AAATTTACAAATCAGTATTCACAAAGTTGCTGTTGCGCAGTGTGAGTCTCAACCTAC	1620
QY	1693	CTTCAATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTTCCCTCTTCTACCAATACCAAT	1752
Db	1621	CTTCAATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTTCCCTCTTCTACCAATACCAAT	1680
QY	1753	TTTGCAAGTACAGTCTACCTCGAAACGATCTAAGATGTGAGTCTTAGTAAGTAACAAAA	1812

Db 1581 TCTGAGTACAGTCTACTCGAA CGCATCTAGAGTGCAGTTTCTAGTAAAGTACAAAA 1740
Qy 1813 CCGATCCCCCAGTGAATCTCTGCTCCAGCCGCGTGAATGGAATCCAAAGCTGAAC 1872
Db 1741 CCGATCCCCCAGTGAATCTCTGCTCCAGCCGCGTGAATGGAATCCAAAGCTGAAC 1800
Qy 1873 TCCAGCGTGTGCTGCTGCTATGTCGCGCCGAGTCCCTCTGAGGACTCTGACGAGAGTCAAG 1932
Db 1801 TCCAGCGTGTGCTGCTGCTATGTCGCGCCGAGTCCCTCTGAGGACTCTGACGAGAGTCAAG 1860
Qy 1933 GGGCTGGCAAGAGGAATGGATTTGGTACGATTTGTGAGCTCCCACTCTCCCGGCCAAGT 1992
Db 1861 GGGCTGGCAAGAGGAATGGATTTGGTACGATTTGTGAGCTCCCACTCTCCCGGCCAAGT 1920
Qy 1993 GCCGAAGTATGAGGCGCACTCCCGCACGAGTTCGAAGAACCCCATGACCCCTAAACGGTGT 2052
Db 1921 GCCGAAGTATGAGGCGCACTCCCGCACGAGTTCGAAGAACCCCATGACCCCTAAACGGTGT 1980
Qy 2053 AATAGTGCAGACGACGAGTGAACCCGGAAGAACGGCCCTAGCGCTGATGGTCCGAGC 2112
Db 1981 AATAGTGCAGACGACGAGTGAACCCGGAAGAACGGCCCTAGCGCTGATGGTCCGAGC 2040
Qy 2113 TGCCAAAGCCAGCTGCGCTGCACCTCAGAAAATCCCTTTGCTAAGGCCAAACGGTCTTCT 2172
Db 2041 TGCCAAAGCCAGCTGCGCTGCACCTCAGAAAATCCCTTTGCTAAGGCCAAACGGTCTTCT 2100
Qy 2173 GGAAGTTGATGCTGCTGCTTGTGCTCTCTCCAGAGACAAAAATCTTAGAGACCTTC 2232
Db 2101 GGAAGTTGATGCTGCTGCTTGTGCTCTCTCCAGAGACAAAAATCTTAGAGACCTTC 2160
Qy 2233 AGGCTTAGCAACAACCTGAAGGCTCGCGGATGAATGAGTGACCTGGACGAGAGG 2292
Db 2161 AGGCTTAGCAACAACCTGAAGGCTCGCGGATGAATGAGTGACCTGGACGAGAGG 2220
Qy 2293 GGCCCTCCGAGGACCGGACCGCGAGCTCAGCTGCGAGCCCGCCGCCGAAATCCCTG 2352
Db 2221 GGCCCTCCGAGGACCGGACCGCGAGCTCAGCTGCGAGCCCGCCGCCGAAATCCCTG 2280
Qy 2353 GAGGACCGAGATGCGGCGCGCAAGCTT 2378
Db 2281 GAGGACCGAGATGCGGCGCGCGCT 2306

RESULT 5

US-10-379-981-11
; Sequence 11, Application US/10379981
; Publication No. US20040001820A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharmaceuticals, Inc.
; APPLICANT: Hahn, Chang S
; APPLICANT: Liu, Hong S
; TITLE OF INVENTION: HUMAN DEUBIQUITINATING PROTEASE GENE ON CHROMOSOME 7 AND ITS
; FILE OF INVENTION: MURINE ORTHOLOG
; FILE REFERENCE: USAV2002-0022 WO PCT
; CURRENT APPLICATION NUMBER: US/10/379,981
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: GB 0218518.9
; PRIOR FILING DATE: 2002-09-08
; PRIOR APPLICATION NUMBER: US 60/366,601
; PRIOR FILING DATE: 2002-03-22
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 3981
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-379-981-11

Query Match 61.7%; Score 1487.6; DB 16; Length 3981;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 419; Indels 51; Gaps 6;
Qy 73 ATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGATCAGCCT 132

Db 1 ATGACCATAGTTGACAAA---ACTGAACCTTCAGACCCATCAACCTGTTCAGAACAGCCT 57
Qy 133 GCGAGCTCCGAGGAGTCTCACCTGGAGACATGAGATGAGGTTCTGCGCAGCTGGGGTGT 192
Db 58 GCGAGTTGTGAGGGCGTCTCACCTGAAGACATGGACACAGAGGCTCTGCCAGCTGGGGCGT 117
Qy 193 GTGTCTTCATTGAATGATGTCCTCAATCAGACACCTTTCTTTAGGACACAGTACTCTGTGT 252
Db 118 GTGTCTTCATTAAGTGTCTCAAGTCAACACTTCCATTAGGGCCAGTGTCTGTGT 177
Qy 253 GTAGTTTATTCGAGTTCATCTGTAACCTATAAATCAAAAACCATCACCAAAAAGATCAA 312
Db 178 GTAGTTTATTCGAGTTCATCTGTAACCTGTAACCTGTAACCTGTAACCTGTAACCTGTA 237
Qy 313 GCGCTAGTGTGATGCGATCGCTCTCCCAAGAAAGTCTTTTCCCACTGAGAGAAATTTGT 372
Db 238 GTCTCTAGTGTGATGCGATCTCTCTCAAAAGGTCCTGTTTCCCATCTGAAAGATTTGT 297
Qy 373 CTTAAGTGGCAACAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 432
Db 298 CTTAAGTGGCAACAAAGTCTAGAGTTGGCGCTGGGCTCCAGAAATTTGGGCAACCTGT 357
Qy 433 TTTGCAATGCGACACTGCAGTGTTTAACTTACCTACACCAACCTCTTGGCAATTTACATGCTA 492
Db 358 TTTGCAATGCGGATTTGCGAGTGTGACTTACAGCCACCCCTCGCCCAATTTACATGTTA 417
Qy 493 TCACATGAACACTCCAAACATGTCATGCAAGAGCTTTTGTATGATGTGTAACAAGTAA 552
Db 418 TCCCATGAACACTCCCAAGACATGCCAGCAGAGGATTTTGTATGATGTGCAAGTGCAG 477
Qy 553 GCACATATACCCAGGCACTCAGTAATCTCGGGACGTTTATAACCAATGTTTGTCTATC 612
Db 478 ACACACATTAACCCAGGCACTTAGCAACCTCGGGATGTTATCAAGCCGATGTTCTGCTATC 537
Qy 613 AATGAGATGCGCGTATAGCTAGGCACCTCCGTTTGGAAACCAAGAAAGATGCCCATGAA 672
Db 538 AATGAAATGCGCGTATAGCTAGACACTTCGTTTGGAAACCAAGAAAGATGCCCATGAA 597
Qy 673 TTCCTTCAATACACTGTTGATGCTATGCAAGAGCATGCTGAATGGCAGCAATAAATTA 732
Db 598 TTTCTTCAGTACACGGTGCATGCCATGCAAGAGCATGTTTAAATGGCAGCAATAAATTA 657
Qy 733 GACAGACACCCAGGCCACCACTCTGTTGTCTAGATATTTGGAGGATACCTTAAGATCT 792
Db 658 GACAGACACCCAGGCCACCACTCTGTTCTCCAGATATTTGGAGGTACCTTAAGATCC 717
Qy 793 AGAGTCAATGTTTAAATTTGCAAGGGCGTTTCAGATACCTTTTGATCCCATATCTTGATATA 852
Db 718 CGAGTTAAATGTTTAAATTTGCAAGGGTGTTCAGATACCTTTGATCCCATATCTGGACATA 777
Qy 853 ACATTTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCAATGGAGAGTTGTGAAGCG 912
Db 778 ACGTTGGAGATTAAAGCTGCACAGAGTGTTCACCAAGGGCTTAGAGCAGTTCGTGAAGCCA 837
Qy 913 GAACAGCTGTGAGAGAAACTCGTACAGTGCAGCAGTGTAAAGAGATGTTCCAGCT 972
Db 838 GAACAACTGGATGGAGAAACTCTACAGTGCAGAGTGCAGAAAGTGCAGAAAGATGTTCCAGCT 897
Qy 973 TCAAGAGAGGTTCACTATCCATAGATCTCTAATGTTTCTTACACTTTCTCTGAAAACGTTTT 1032
Db 898 TCAAGAGATTCACAAATCCATAGGTCCTCTAATGTTTCTTACCATCTCACTGAAGCGCTTT 957
Qy 1033 GCAATTTTACCGGTGGAAAAATTCCTAAGGATGGAATACCCCTGAGTATCTTGATATT 1092
Db 958 GCCAACTTCACCGGTGGAAAGATTCCTAAGGATGTGAATATCTCTGAGTACCTTGATATC 1017
Qy 1093 CGGCCATATATGCTCAACCCCAAGGAGGCAATTTGCTAGCTTTGTATCGAGTGTG 1152
Db 1018 CGGCCCATATATGCTCAGCCCAATGAGAGCAATTTATTTATGTTTGTATGCTGTGCTG 1077
Qy 1153 GTCCACACTGTTTAAATTTGCCATGCTGGCCATTTACTTCTGCTACATAAAGCTAGCAAT 1212

Db 1078 GTGCACACTGGTTTAAATTTGCTATGCTGGCCACTACTTTTGTCTACATCAAGGCTAGCAAT 1137
QY 1213 GGCTCTGGTATCAAAATGAATGACTTCATTTGTATCTACAGTATATATAGATCGTACTC 1272
Db 1138 GGCTCTGGTATCAGATGAATGATCTCATCGTGTCCACCAAGTATATCAGAGCAGTCTT 1197
QY 1273 AGCCAAACAGCTATGTGCTCTTTTATATATCAGGTGCCATGATGTGAAAAATGAGGTGAA 1332
Db 1198 AACAGCAAGCTTACGTGCTCTTTTATATATCAGTCCCTATGATGAAAAATGAGGGGAG 1257
QY 1333 CTATCTATCCCAACCCATAGCCCGGCGGAGTCTCTCCCGCCCGGCTCATCAGTCAGCGG 1392
Db 1258 TCTGCTCATCTGCTCCATAGCCCGGCGCAATCTCTCCCGCCCGGAGTCAAGTCAGCGG 1317
QY 1393 GTTGTACCAACAAACAGGCTGCGCAGGCTTTATCGGACACACAGTTCCTCTCACATG 1452
Db 1318 GTAGTCAACAAACAGGAGGTGCTCCAGGGTTTATTGGACCCCGAGCTGCTTCCCATGTG 1377
QY 1453 ATAAAGAAATCCACTCACTTAAATGGGACTGGACCAATTGAAAGACACGCCAAGCAGTTCC 1512
Db 1378 ATGAAGAACAGCCACACACTTGAATGGCACCACCGCCAGTGAAGACACACCAAGTAGTTCT 1437
QY 1513 ATGTGAGTCTTAAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTGTAATGTTTCAAGT 1572
Db 1438 GTCTCAAGCCCTTAACCGAAACACACAGCGTCAATAGGGCCAGTCTGTGCTTGGACT 1497
QY 1573 TCTGTCCAAACCTGCTCAGTAAATAGTCTCTCAGTGATCCAGAAATCCTTAAAGAACAA 1632
Db 1498 TCTGTGAGAACTGGTCTGTGTTACAGACCCCTCAGTTATTCAGATCACCCCAAGAAACAA 1557
QY 1633 AAAATTAACATCAGTATTCACAAACAGTTGCCCTGTGTCGCCAGTGTCACTCTCAACCTAA- 1691
Db 1558 AAAATCACCATCAGTATTCACAAACAGTTGCCCTGTGTCGCCAGGTGAGGACCACTGAAT 1617
QY 1692 -----CCTTCATAGTAAATCTTTGGAGAACCCCTACCAAGCCGCTGCTCTTCTTACAAAT 1746
Db 1618 AACAGGCTCCATGSCCTTGTCTGGAGGCTCTCTAGTAAGCGCGCACCCCTCTCCACCATC 1677
QY 1747 ACCAA---TTCGTGACGTACAGTCTACTCGAAGCATCTACGATGTCAAGTTTCTAGTAA 1803
Db 1678 ACTAACCTTCTCGAATACAGTCTACTCTCGAAGCATCTACCCCAACGTT----- 1723
QY 1804 GTAACAAAACCGATCCCGCAGTGAATCCTGTCTCCAGCCCGCTGATGAATGCCAAATCC 1863
Db 1724 -----CGACTTCCCGCAGTGAAGCTGTGTCCTCCCAAGCCATGTTGAACGCGAGCT 1773
QY 1864 AAGCTGAATCAGCGTGTGTCCTATGCGCGCGAGTCTCTCGAGGACTCTGACGAG 1923
Db 1774 AAGTGGCGCCAGTGTGCTGTGCTTATGGCGCGAGTCTCTCAGAAGAGTCTGATGAG 1833
QY 1924 GAGTCAAAAGGGCTGGCAAGGAGATGGATTTGTTACGATTTGTAGCTCCCACTCTCC 1983
Db 1834 GAGTCAAGGGCTGGCAAGGAGAACGGTGTGGAATGATGGCGCGCACTCACTCCGAT 1893
QY 1984 GGCCAAAGATGC-----CGAAGATGAGGAGGCCACTCCGACAGGCTTCAAGAACCC 2034
Db 1894 AGGCCAAGAGCTGCTGCAGATGACGCTGTGAGGCTTCTCCCATGAGCTTCAAGAACCC 1953
QY 2035 ATGACCCCTAAACGGTCTTAATAGTGCAGACAGGACAGTGAACCCGAAAGAAACGGCCTA 2094
Db 1954 GTCTCTTAATAGTGTCTAATAGCGCA-----GACAGTGACTTCAAGAGAAACAGCCTG 2007
QY 2095 GCGCCTGATGGTCCAGCTGCCAAGGCCAGCCTGCGCTGCACTCAGAAAATCCCTTTGCT 2154
Db 2008 GCATTTGACAGTGCAGCTGCCAGGTCCAGCCCGAGCTACACAGAAACCTCTTTTCC 2067
QY 2155 AAGGCAAAACGGTCTTCTCTGAAAGTTGATGCCCTGCTCTTTTGTCTCTCCAGAAAGAC 2214
Db 2068 AAATTTAAATGGTCTTCTCTGAAAGTGAAGCGCTGCTCTTGTGAGTCTGTCTCTGAGAC 2127
QY 2215 AAATCTTACAGACCTTACAGCTTACCAAACTGAAAGGCTCGACGGATGAATGAGT 2274
Db 2128 AGAATCCTTGAGACCTTCAAGCTTACCAACCGAGGCAAGGGTCCAGCGGGTGAAGAGT 2187

QY 2275 GCACCTGGACAGAGAGGGGCGCTCCCGAGGACGGGACCGCGAGCCTCAGCTGGCAGC 2334
Db 2188 TGGACTACGACAGAGGGGGAAGCTCTCCAAAGGACCCCTGTTTTCACAGCTGGAGCCCATCAGT 2247
QY 2335 CCGCGCCGCCAAATCCCTGGAGGAGCAGATGCGCGCCGCAA 2374
Db 2248 GATGAGCCCAAGTCCCTTGGATACCGGAGCTGTACCA 2287

RESULT 6

US-10-371-905A-33
; Sequence 33, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: human DUB8.7
US-10-371-905A-33

Query Match 18.4%; Score 443; DB 13; Length 1593;

Best Local Similarity 62.7%; Pred. No. 8e-132;
Matches 733; Conservative 0; Mismatches 415; Indels 21; Gaps 2;

QY 163 ATGGATGCAGGTTCTGCCAGCTGGGTGCTGTCTTTCAATTGAATGATGTCAAAATCAC 222
Db 1 ATGAGGACGACTCACTCTACTTGGAGGTGAGTGGCAGTTCAACCACTTTCAAACATC 60
QY 223 ACATTTCTTTAGGACCACTGCTGTGTGTGTATTTATTCGAGTTTCATCTGACCTGAT 282
Db 61 ACATCTTCTGGCCAGATGACAGCTTTTGTGTAATCCA---GGGACTTCTCTCCCTGAG 117
QY 283 AATCAAAACCATCAACAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAG 342
Db 118 AAGTACCACCTCTCATCTGAGGCCCGTGTGCACCTCTGTGATGATTTGGCTCTCTGGCA 177
QY 343 AAAGTTCTTTTCCCATCTGAGAAGATTGTCTTTAAGTGGCAACAACTCATAGAGTTGGA 402
Db 178 AGACAGCTTGTCTCCAGGAGAGGCTTCTCTGAGTAGAGGAGACCTGCTGCGGTGGG 237
QY 403 GCTGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATGACAGCACTGCAAGTGTTTAACC 462
Db 238 GCTGGGCTCCAGAAATATGGAAATACCTGCTACGAAACGCTTCTCTGCAAGTGCCTGACA 297
QY 463 TACACACCACTCTTCCCAATTCATGCTATCATGAACACTCCAAACATGTCATGCA 522
Db 298 TACACACCCCGCTTCCCAACTACATGCTGTCCGGGAGCACTCTCAAAATGTCAGCGT 357
QY 523 GAAGGCTTTGTATGATGTGTACAATGCAAGACATATTACCCAGGCACCTCAGTAATCTCT 582
Db 358 CCCAAGTGTGCATGCTCTGTACTATGCAAGCTCACATCACATGGGCGCTCCACAGTCTCT 417
QY 583 GGGAGCTTATTAAACAAATGTTTGTCAATGATGGGGGTATAGTAGGCACTTC 642
Db 418 GGTCA-----TGTCACTCCAGCCCTCAAGGCATTTGGCTGTGGCTTC 459
QY 643 CGTTTTGGAAACCAAAGAGATGCCCATGAATTCCTTCAATACATCTGTGATGTCATGCAG 702
Db 460 CATAGAGGACGAGAGAGATGCCCATGAATTTCTCATGTTCTACTGTGATGATGCAAAA 519
QY 703 AAAGCATGCTTGAATTCGACAGCAATAAATTAGACAGACACACCCAGGCCACCACTTTGTT 762

Db 520 AAGGATGCTTCCGGCCCAAGAGGTAGATCATCTAAGGACACACCCCTATC 579
Qy 763 TGTGAGATATTTGGAGGATACCTAAGATCTAGATCAAAATGTTTAAATTTGAAGGGGCTT 822
Db 580 CACCAAAATATTTGGAGGCTGCTGGAGATCTCAAATCAAGTGTCTCCACTGCGCAGGGATT 639
Qy 823 TCAGATCTTTTGTATCCATATCTTGTATATACATTTGGAGATAAAGGCTGCTCAGAGTGT 882
Db 640 TCAGACACTTTTGTAGCCCTTACCTGGACATCGCCCTGGATATCCAGGCAGCTCAGAGTGT 699
Qy 883 AACAGGCATTGGAGCAGTTTGTGAAGCGGAACAGCTTGTAGTGAAGAAATCTCGTACAA 942
Db 700 AAGCAGCTTTTGGAAACAGTTGTGAAGCCGGAAGAACTCAATGGAGAGAAATGCTTAT 759
Qy 943 TGCAGCAAGTGTAAAGAGTGTCCAGCTTCAAGAGGTTTCACTATCATATAGATCTCT 1002
Db 760 TGGGTCTTTGTCTCCAGAGGGCGCAGCTTCAAGAGCTTAACTTTACACACTTCTGCC 819
Qy 1003 AATGTTCTTACACTTCTCTGAAAGCTTTTGAAGCTTTTACCGTGGAGAAATTTGCTAAG 1062
Db 820 AAGGTCTCTATCTTGTCTTGAAGAGATTTCTCCGATGTACAGGCAACAAACTTGCCTAAG 879
Qy 1063 GATGTGAATACCTGTAGTATCTTATATTCGGCCATATATGTCTCAACCCAAACGGAGAG 1122
Db 880 AATGTGCAATATCTGAGTGCCTTGAATGAGCCATATACATATGTCTCAGCAGACACAGA 939
Qy 1123 CCAATGTCTAGCTTGTATGACGTGTGTCACACTGTTTAAATTTGCAAGTCTGCC 1182
Db 940 CCTCTGTCTATGCTCTATGCTGTGTGTCACGCTGGGTGGAGTTGTCAAGATGGA 999
Qy 1183 CATTACTTCTGCTACATAAAGCTAGCAATGGCTCTGCTATCAATGAATGACTCCATT 1242
Db 1000 CATTACTTCTCTTATGTCAAAGCTCAAGAAGCCAGTGTGTATATAATGGATGCCAAG 1059
Qy 1243 GTATCTACAGTATATAGATCGGTACTCAGCCAAAGCCATGTGTCTTTTATATC 1302
Db 1060 GTCACTGCTGTAGCATCACTTCTGCTGTGATCAACAGGCCATGTCTCTTTTACATC 1119
Qy 1303 AGGTCCCATGTGTGAAATAAGAGGTGA 1331
Db 1120 CAGAAGATGAATGGAAAGACACAGTGA 1148

RESULT 7

US-10-371-905A-19
; Sequence 19, Application US/10371905A
; Publication No. US2003022496A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 1698
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: human DUBA.8
US-10-371-905A-19

Query Match 18.1%; Score 436; DB 13; Length 1698;
Best Local Similarity 62.1%; Pred. No. 1.6e-129;
Matches 732; Conservative 0; Mismatches 425; Indels 21; Gaps 2;
Qy 154 CTGGAGACATGGATGAGGTTCTGCGAGCTGGGGTGTGTCTTCAATTGAATGATGTG 213
Db 97 CCAGTCGACATGGAGGACCACTCACTTACTTGGGAGGTGAGTGGCAGTTCAACCACTTT 156

Qy 214 TCBAATCACACACTTTCTTTAGGACCACTACCTGGTCTGTAGTTTATTCGAGTTCATCT 273
Db 157 TCAAAACTCACATCTTCTCGGCCGATGCAGCTTTTGTGAAATCCA--CGGACTTCT 213
Qy 274 GTACTGTATAAATCAAAACCAATCACCAAAAGGATCAAGCCCTAGAGTGATGGCATCGCT 333
Db 214 CTCCTGAGAGTCAACCACTCTCATGTGAGACCCGTGTGACCTCTGTGATGATTTGGCT 273
Qy 334 CCTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTTGTCTTAAAGTGGCAACAAACTCAT 393
Db 274 CCTGTGCAAGACAGCTTGTCTCCAGGGAGAAAGTTCTCTCTAGTAGCAGGAGACCTGCT 333
Qy 394 AGATTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATGCAGCACTGCGAG 453
Db 334 GCGGTGGGGCTGGCTCCAGAAATATGGGAATACCTCTAGTGAAGCTTCTTCTTGGAG 393
Qy 454 TGTTTAACTCACACACCACTCTTGGCCAAATTACATGTCTATCACATGAACACTTCCAAAA 513
Db 394 TGCCTGACATACACACCGCCCTTGGCAACTACATGCTGTCCCGGAGCACTCTCAAAG 453
Qy 514 TGTATGCAAGAGGCTTTTGTATGATGTACATGCAAGCAATATTAACCCAGCACTC 573
Db 454 TGTATCTGTCAAGGGCTGCAATGCTCTGTAGTCAAGCTCAATCACACGGGCCCTC 513
Qy 574 AGTAATCTCTGGGGAGTTTAAACCAATGTTGTCTCATCAATGAGATGGGGGTATAGCT 633
Db 514 CACAATCTCTGGCCAGCTCATCCAGCC-----CTCAGGCAATTTGGCT 555
Qy 634 AGGCACCTCTCGGTTTGGAAAACCAAGAGATGCCATGAATTCCTTCAATFACACTGTGTAT 693
Db 556 GCTGCTTCCATAGAGGCAAGCAGAGATGCCATGAATTTCTCATGTTCACTGTGGAT 615
Qy 694 GCTATGCAAGAGCATGCTTGAATGGCAGCAATAAATAGACAGACACACCAGGCCACC 753
Db 616 GCCATGAAAAGGCAATGCTTCCCGGCCACAGCAGGTAGATCATCACTCTAAGGACACC 675
Qy 754 ACTCTGTGTTGTGAGATATTTGGAGATACCTTAAGATCTAGAGTCAAAATGTTTAAATGTC 813
Db 676 ACCCTCATCCACCAAAATATTTGGAGGCTACTGAGATCTCAAAATCAAGTGTCTCCACTGC 735
Qy 814 AAGGCGTTTCAGATACATTTTGCATCCATATCTTGATATAACAATTTGGAGATAAAGGCTGT 873
Db 736 CACGCAATTTCAGACACTTTTGACCTTACCTGGACATCGCCCTGGATATCCAGGCAGCT 795
Qy 874 CAGAGTGTCAACAGGCATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGTATGGAGAAAC 933
Db 796 CAGAGTGTCCAGCAAGCTTTGGAAACAGTTGTGTGAAGCCCGAAGAACTCAATGGAGAGAT 855
Qy 934 TCGTACAAAGTGCAGCAAGTGTAAAGAGATGGTTCCAGCTTCAAGAGGTTCACTATCCAT 993
Db 856 GCCTATCATTTGGGTGTTGTTCTCAGAGGGCGCGGCTCCAGACGTTAATTTTACAC 915
Qy 994 AGATCCTCTAATGTTCTTACACTTTTCTCTGAACGTTTTTGAATAATTTACCGTGGAAAA 1053
Db 916 ACCTCTGCCAAGGTCCTCATCTTGTATTGAAGATTTCTCGATGTGACAGGCAACAG 975
Qy 1054 ATTGTAAAGATGTGAATAACCTGAGTATCTTGATATTTGGCCCATATATGTCTCAACC 1113
Db 976 ATTGCCAAGATGTGCAATATCTCTGAGTGCCTTGACATGTCAGCCATACATGTCTCAGCAG 1035
Qy 1114 AACGGAGGACCAATTTGTCTACGTTGTATGAGTGTGCTGCTCCACACTGTTTAAATTCG 1173
Db 1036 AACACAGGACCTCTTGTCTATGTCTCTATGTGTGTTCCAGCTGGGTGGAGTGTGT 1095
Qy 1174 CATGCTGGCCATTACTTCTGTACATAAAGCTAGCAATGGCTCTGGTATCAAAATGAT 1233
Db 1096 CACAAGGACATTAATCTTCTTATGTCAAGCTCAAGAGGCCAATGGTATATAAATGAT 1155
Qy 1234 GACTCCATTTGATATCACAGTGAATTTAGATCGGTACTCAGCCCAACAGCCCTATGTGCTC 1293
Db 1156 GATCGCGAGGTCAACCGCGCTAGCATCTTCTGTCTTGAAGTCAACAGGCCCTACGTCCTC 1215

Qy 1294 TTTTATATCAGGTCCCATGATGTGAAAAATGGAGGTGA 1331
||||| ||| | ||| | |||
Db 1216 TTTTATCATCCGAAAGAGTGAATGGGAAAGACACAGTGA 1253

```

RESULT 8
US-10-107-695B-1
; Sequence 1, Application US/10107695B
; Publication No. US20030022201A1
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 68999, Human Ubiquitin
; FILE REFERENCE: MPIO1-076PIRCPIM
; CURRENT APPLICATION NUMBER: US/10/107,695B
; CURRENT FILING DATE: 2002-09-10
; PRIOR APPLICATION NUMBER: 60/279,184
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1763
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (171) ... (1763)
US-10-107-695B-1

```

Query Match	18.0%;	Score 434.6;	DB 15;	Length 1763;
Best Local Similarity	62.0%;	Prod. No. 4.6e-129;		
Matches 733;	Conservative	0;	Mismatches 429;	Indels 21; Gaps 2
QY	149	TCTCACCTGGAGACATGATCGAGCTTCTGCGAGCTGGGGTGTCTGTGCTTCATTTGAATG	208	
Db	157	TCGTTCCAGTCGACATGGAGGACGACTCACTCTACTTTGGGAGGTGAGTGGCGAGTTCAACC	216	
QY	209	ATGTGTCAAATCACACACTTTCTTTTAGGACCAAGTACCTGGTGTCTGTAGTATTTATTCGAGTT	268	
Db	217	ACTTTTCAAAACTCACACTCTTCTCGGCCGATGACGCTTTTGTCTGAATCCA--GGCGA	273	
QY	269	CATCTGTACTGATATAATCAAAACCATCACACAAGGATCAAGCCCTAGTGTATGCGA	328	
Db	274	CTTCTCTCCCTGGAAGTCACCATCTCATGTGAGGCCGCTGTGACCTCTGTGTATGATT	333	
QY	329	TCGCTCTCTCCACAGAAAAGTTCTTTTCCCATCTGAGAAATTTGTCTTAAGTGGCAACAA	388	
Db	334	TGGCTCTCTGGCAAGACAGCTTCTCCAGGGAGAGCTTCCCTGTAGTAGCAGGAC	393	
QY	389	CTCATAGAGTTGAGCTGGGCTCAGAAATTTGGGCAATACCTGTTTTCGCAATGAGCAGC	448	
Db	394	CTGCTGGGTGGGGGCTGGGCTCCAGAAATATGGGAAATACCTGCTACGTGTAACGCTTCT	453	
QY	449	TGCAGTCTTTAAACCTACACACCACTCTTGGCAATTCATCTCATCATGAACACTCCA	508	
Db	454	TGCAGTGCCTGACATACACACCGCCCTTGGCAACTCATGCTGTCCGGGAGCACTCTC	513	
QY	509	AAACATGTCTATGAGAGGCTTTTGTATGATGTGTACAAATGCAAGCACATATTACCCAGG	568	
Db	514	AAACGTTGCATCGTCAACAGGGCTGCATGCTCTGTACTATGCAAGCTCACATCAACGGG	573	
QY	569	CACCTAGTATCTGGGAGCGTTATTAAACCAATGTTTGTTCATCAATGAGATGGCGGTA	628	
Db	574	CCCTCCAAATCTCTGGCCAC-----GTCAATCAAGCCCTCAACAGGCAT	615	
QY	629	TAGCTAGGCACCTCCGTTTGTGGAAACCAAGAATGCCCATTGAAATTCCTTCAATACACTG	688	
Db	616	TGGCTGTGGCTTCCATAGAGGCAAGCAGGAAGATGCCATGAATTTCTCATGTTCACTG	675	
QY	689	TTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATTAATTAGACAGACACCCAGG	748	
Db	676	TGATGCCATGAAAGAGCATGCTTCCGGGCAACAGCAGGTAGATCATCACTCTAAGG	735	

Qy	749	CCACC	CTCTGTTGTG	CAGATATTTGG	AGGATACCTAAG	ATCTAGAGTCAAA	TGTTAA	808
Db	736	ACACC	CCCTCATCC	ACAAATA	TTGGAGGCT	ACTGAGATCTCAAA	TCAAGTCTCC	795
Qy	809	ATTCC	AAGGGGTTT	CAGATAC	TTTTGATCCATAT	CTTTGATATAAC	CATTGGAGATAA	868
Db	796	ACTCC	ACGGCATTT	CAGACATTTG	ACCCCTTACCTTGG	ACATCGCCTGGATAT	TCCAGG	855
Qy	869	CTGCT	CAGAGTGTCA	ACAGSCATT	GGACAGTTTGTGA	AGCCGGAACAGCT	TGATGGAG	928
Db	856	CAGCT	CAGAGTGT	CAGCAAGCTTT	GAAACAGTTGGTGA	AGCCCGAAGAACTCAAT	GGAG	915
Qy	929	AAAA	CTCGTACAAGT	GCAGCAAGTGTAAA	AGATGGTTC	CAGCTTCAAAGAGTTCACTA		988
Db	916	AGAAT	GCTATCAT	TGTGGTGT	TTGTCTCAGAGGGCG	CGGCTCCAAGACGTTAACTT		975
Qy	989	TCCAT	AGATCCTCTA	ATGTTCTTACACTTT	CTGAAACGTTTTT	GTGCAAAATTTTACCGGTG		1048
Db	976	TACAC	ACTCTGCCAAGGT	CTCATCTTGATTTGA	AGAGATTTCTCGATGT	TCAGGCA		1039
Qy	1049	GA	AAAAATTGCTA	AGCATGTGAAT	TACCTCAGTATCTTGATAT	TCGGCCATATATGCTC		1108
Db	1036	ACAAG	ATTGCCAAGAT	GTCAATATCCT	CAGTGCCTTGACAT	CGACCCATACATGCTC		1099
Qy	1109	A	CCCCAGGAGAGC	CAATTGTCTAC	GTCTTTGTATG	CAGTGTCTGGTCCACACTGGTTTTA		1168
Db	1096	AGCAG	AAACAGAGAC	CTTTGTCTATGTCCT	TATGTCCTCTATGCTGTG	TGTCAGCTGGTCCACGCTGGGTGA		1155
Qy	1169	ATTGCC	ATGTCGGCATTA	CTTCGTACATATAA	AGCTAGCAATGGCTCTGGTAT	CAAA		1228
Db	1156	GTGT	GCACAGGACATTA	CTTCTCTTATGTCA	AAAGCTCAAGAGGCGAGTGGTAT	AAAA		1215
Qy	1229	TGAAT	GACTCCATTTGAT	CTACCAGTGATATTAGAT	CGGTACTCAGGCCAA	CAAGCCCTATG		1288
Db	1216	TGGAT	GATGCGAGGT	CACCGCTCTAGCAT	CACTTCTGTCTTGAGTCA	CAGGCCCTACG		1275
Qy	1289	TGCT	CTTTTATATCAG	GTCCCATGATGTG	AAAAATGGAGGTGA			1331
Db	1276	TCCT	CTTTTATATCC	CAAGAGGTGAAT	TGGGAAACACAGTGA			1318

```

RESULT 9
US-10-371-905A-17
; Sequence 17, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang
; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: USAV2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: human DUB4.7
US-10-371-905A-17

```

	Query Match	18.0%	Score 433.4	DB 13	Length 1593
	Best Local Similarity	62.3%	Pred. No. 1e-128		
	Matches 727	Conservative 0	Mismatches 421	Indels 21	Gaps 2
QY	163	ATGGATCGAGGTTCTGCCAGCTGGGGGTCGTGCTTCATTGAATGATGTCGTCAAAATCAC	222		
Db	1	ATGGAGGACGACTCATCTCTACTTGGGTGGTGAGTGGCGAGTTCAACCACCTTTTCAAAATC	60		
QY	223	ACACTTCTTTTAGGACGAGTACCTGGTGCCTAGCTTTATTTCGAGTTCACTGTACCTGAT	282		

Db 61 ACATCTTCTCGCCGATGACGCTTTTGTGTAATCCA---CGGACTTCTCTCCTCGAG 117

QY 283 AAATCAAAACCATACCAAAAGGATCAAGCCCTGAGTGATGGCATCGCTCTCCACAG 342

Db 118 AAGTCACCACTCTCATGTGAGACCGCTGCGACCTCTGTGATGATTTGGCTCTCTGGCA 177

QY 343 AAAGTTCTTTTCCCATCTGAGAGATTTGTCTTAAGTGGCAACAACCTCATAGAGTTGA 402

Db 178 AGACAGCTGTCTCCAGGGAGAGCTTCTCTGAGTAGCAGGACCTGCTCGGTGGG 237

QY 403 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACTGCGAGTTTAAAC 462

Db 238 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACTGCGAGTTTAAAC 297

QY 463 TACACACCACTCTTGGCAATTTACATGTTATCACAATGAACATCCAAAACATGTCATGCA 522

Db 298 TACACACCGCCCTTGCCAGGGAGAGCTTCTCTGAGTAGCAGGACCTCTCAAAACGTGTCATG 357

QY 523 GAAGCTTTTGTATGATGTGTAATGCGCAATTTAGACAGACACACCGGCCACCACTCTTGT 582

Db 358 CACAAGGCTGATGCTCTGTGACGATGCAAGCTCACAATCACACGGGCTCTCCACATCT 417

QY 583 GGGAGCGTTATTAACCAATGTTTGTCTATCAATGATGCGCGTATAGTAGGCACTTC 642

Db 418 GGCCACGTCATCCAGCC-----CTCACAGGCATTTGGCTCTGGCTTC 459

QY 643 CGTTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACATCTGTTGATGCTATGCA 702

Db 460 CATAGAGGCAAGCAAGATGCCCATGAATTTCTATGTTCCATGTTGATGATGCCATGAA 519

QY 703 AAAGCATGCTTGAATGCGCAATTTAGACAGACACACCGGCCACCACTCTTGT 762

Db 520 AAGGCATGCTTCCCGGCCACAGCAGGTAGATCATCTCTAAGACACCACTCTATC 579

QY 763 TGTGATATTTGGAGGATCACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTT 822

Db 580 CACCAATATTTGGAGGCTACTGGAGATCTCAATCACTGTCTCCACTGCCACGGCAT 539

QY 823 TCAGATATTTTGTATTCATTTGATATTAACATTTGAGATTAAGGCTGCTCAGAGTGT 882

Db 640 TCAGACATTTTGACCTTTACCTGACATCGCCCTGGATATCCAGGACGCTCAGAGTGT 699

QY 883 AACAGGATTTGGACGCTTTGTGAGCGGACACGCTTGATGGAGAACTCTGTCACAG 942

Db 700 CAGCAAGCTTTGGACGCTTTGTGAGCGGACCAAGCTCAATGAGAGATGCTTATCAT 759

QY 943 TGCAGCAAGTGTAAAGATGCTTCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCT 1002

Db 760 TGTGTGTTTGTCTCCAGAGGGCGCGCTCCAAGACGTTAACTTTACACACCTCTGCC 819

QY 1003 AATGTTCTTACACTTCTCTGAAAAGTTTGGCAATTTTACCGGTGGAAAAATTTGCTAAG 1062

Db 820 AAGGTCCTCATCTTGTATTTGAAGATTTCTCCGATGTCAAGGCAACAAGATTTGCAAG 879

QY 1063 GATGTGAATACCTTGATCTGATATTCGTCATATTCGTCATATCTCAACCCACCGAGAG 1122

Db 880 AATGTGCAATATCTTGAGTGCCTTGACATGCGAGCCATACATGTTCTCAGCAGAACACAG 939

QY 1123 CCAATTTGTCTAGCTTGTATGAGTGTGTTCCACACTGTTTAAATGGCAATGTCGTC 1182

Db 940 CTTCTTGTATGCTCTATGCTGTGCTGTTCCACGCTGGGTGGAGTTGTTCACACGGA 999

QY 1183 CATTAATCTTGTACATAAAGCTAGCAATGGCTCTGTTATCAATGAATGACTCCATT 1242

Db 1000 CATTAATCTTGTATGTCAAAGCTCAAGAGGCGCAGTGGTATATAAATGATGATGCCGAG 1059

QY 1243 GTATCTACAGTATATAGATGCTTACCGCACAGGCTATGCTGCTCTTTTATATC 1302

Db 1060 GTACCGCGCTAGCATCACTTGTGCTGAGTCAACAGGCGCTAGCTCTCTTTTACATC 1119

QY 1303 AGGTCCCATGATGTGAAAAATGGAGTGA 1331

Db 1120 CAGAGAGTGAATGGGAAAGACACAGTGA 1148

RESULT 10

US-10-371-905A-11

; Sequence 11, Application US/10371905A

; Publication No. US20030224969A1

; GENERAL INFORMATION:

; APPLICANT: Hahn, Chang

; APPLICANT: Liu, Hong

; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes

; FILE REFERENCE: USAV2001/0185

; CURRENT APPLICATION NUMBER: US/10/371,905A

; CURRENT FILING DATE: 2003-02-20

; NUMBER OF SEQ ID NOS: 54

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 11

; LENGTH: 1593

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; OTHER INFORMATION: human DUB4.3

US-10-371-905A-11

Query Match 17.9%; Score 431.8; DB 13; Length 1593;

Best Local Similarity 62.1%; Pred. No. 3.5e-128;

Matches 726; Conservative 0; Mismatches 422; Indels 21; Gaps 2;

QY 163 ATGGATGACAGGTTCTGCGCAGCTGGGGTGTGTCTTTCATTTGAATGATGTGCAATCAC 222

Db 1 ATGGAGGACGACTCACTTACTTGGAGGTGAGTGGCAGTTCACCACTTTTCAAACTC 60

QY 223 ACATCTTCTTTAGACACGACTGCTGGTGTGTGATTTTATTCAGTTCATCTGTACTGAT 282

Db 61 ACATCTTCTCGGCCGATGACGCTTTTGTGTAATCCA---GCGGACTTCTCTCCTGAG 117

QY 283 AAATCAAAACCATACCAACAAGGATCAAGCCCTAGTGTGATGGCATCGCTCTCCACAG 342

Db 118 AAGTCACCACTCTCATGTGAGACCGCTGTCGACCTCTGTGATGATTTGGCTCTCTGGCA 177

QY 343 AAAGTTCTTTTCCCATCTGAGAGATTTGTCTTAAGTGGCAACAACCTCATAGAGTTGA 402

Db 178 AGACAGCTTGTCTCCAGGGAGAGCTTCTCTGATGACAGGACCTGCTCGGTGGG 237

QY 403 GCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACTGCGAGTTTAAAC 462

Db 238 GCTGGCTCCAGAAATTTGGCAATACCTGCTACGTGAAACGTTCTTTCAGTGCCTGACA 297

QY 463 TACACACCACTTGTGCAATTTAGACAGACATATTTACCCAGGCACTCAGTAACTCT 582

Db 298 TACACACCGCCCTTGCCAGGATGATGCTGTCGCGGAGCACTCTCAACGCTGTCATG 357

QY 523 GAAGCTTTTGTATGATGTGTAATGCGCAATTTAGACAGACACACCGGCCACCACTCTTGT 582

Db 358 CACAAGGCTGATGCTCTGTGACGATGCAAGCTCACAATCACACGGGCTCTCCACATCT 417

QY 583 GGGAGCGTTATTAACCAATGTTTGTCTATCAATGATGCGCGTATAGTAGGCACTTC 642

Db 418 GGCCACGTCATCCAGCC-----CTCACAGGCATTTGGCTCTGGCTTC 459

QY 643 CGTTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACATCTGTTGATGCTATGCA 702

Db 460 CATAGAGGCAAGCAAGATGCCCATGAATTTCTATGTTCCATGTTGATGATGCCATGAA 519

QY 703 AAAGCATGCTTGAATGCGCAATTTAGACAGACACACCGGCCACCACTCTTGT 762

Db 520 AAGGCATGCTTCCCGGCCACAGCAGGTAGATCATCTCTAAGACACCACTCTATC 579

QY 763 TGTGATATTTGGAGGATCACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTT 822

Db 580 CACCAATATTTGGAGGCTACTGGAGATCTCAATCACTGTCTCCACTGCCACGGCAT 539

QY 823 TCAGATATTTTGTATTCATTTGATATTAACATTTGAGATTAAGGCTGCTCAGAGTGT 882

Db 640 TCAGACATTTTGACCTTTACCTGACATCGCCCTGGATATCCAGGACGCTCAGAGTGT 699

QY 883 AACAGGATTTGGACGCTTTGTGAGCGGACACGCTTGATGGAGAACTCTGTCACAG 942

Db 700 CAGCAAGCTTTGGACGCTTTGTGAGCGGACCAAGCTCAATGAGAGATGCTTATCAT 759

QY 943 TGCAGCAAGTGTAAAGATGCTTCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCT 1002

Db 760 TGTGTGTTTGTCTCCAGAGGGCGCGCTCCAAGACGTTAACTTTACACACCTCTGCC 819

QY 1003 AATGTTCTTACACTTCTCTGAAAAGTTTGGCAATTTTACCGGTGGAAAAATTTGCTAAG 1062

Db 820 AAGGTCCTCATCTTGTATTTGAAGATTTCTCCGATGTCAAGGCAACAAGATTTGCAAG 879

QY 1063 GATGTGAATACCTTGATCTGATATTCGTCATATTCGTCATATCTCAACCCACCGAGAG 1122

Db 880 AATGTGCAATATCTTGAGTGCCTTGACATGCGAGCCATACATGTTCTCAGCAGAACACAG 939

QY 1123 CCAATTTGTCTAGCTTGTATGAGTGTGTTCCACACTGTTTAAATGGCAATGTCGTC 1182

Db 940 CTTCTTGTATGCTCTATGCTGTGCTGTTCCACGCTGGGTGGAGTTGTTCACACGGA 999

QY 1183 CATTAATCTTGTACATAAAGCTAGCAATGGCTCTGTTATCAATGAATGACTCCATT 1242

Db 1000 CATTAATCTTGTATGTCAAAGCTCAAGAGGCGCAGTGGTATATAAATGATGATGCCGAG 1059

QY 1243 GTATCTACAGTATATAGATGCTTACCGCACAGGCTATGCTGCTCTTTTATATC 1302

Db 1060 GTACCGCGCTAGCATCACTTGTGCTGAGTCAACAGGCGCTAGCTCTCTTTTACATC 1119

QY 1303 AGGTCCCATGATGTGAAAAATGGAGTGA 1331

Db 1120 CAGAGAGTGAATGGGAAAGACACAGTGA 1148

Db 178 AGACAGCTTGTCTCCAGGAGAAAGCTTCTCTCAGTAGCAGGAGACCTGCTCGGTGGG 237
QY 403 GCTGGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGCAGACACTGCAAGTTTAAAC 462
Db 238 GCTGGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGCAGACACTGCAAGTTTAAAC 237
QY 463 TACACACCACTCTGTCGAATTAATCATGTAACCACTTCCGAGGAGCACTCTCAAGCTGTCATCGT 522
Db 298 TACACACCGCCCTTGGCAATTAATCATGTAACCACTTCCGAGGAGCACTCTCAAGCTGTCATCGT 357
QY 523 GAAGGCTTTTGTATGATGTGTACAAATGCAAGCAATATTAATCCAGGCACTTCAATATCCT 582
Db 358 CACAAGGCTGTATGCTCTGATGATGCAAGCTTCAATATTAATCCAGGCACTTCAATATCCT 417
QY 583 GGGGAGGTTATTAACCAATGTTTGTATGATGCAAGCTTCCGAGGAGCACTCTCAAGCTGTCATCGT 642
Db 418 GGCACGTCATCCAGC-----CTCACAGGCACTTGGCTGCTGCTTC 459
QY 643 CGTTTTTGGAAACCAAGATGCCATGAATTCCTTCAATACACTGTTGATGCTATGCGAG 702
Db 460 CATAGAGGCAAGCAGAGAGATGCCATGAATTTCTCATGTTCACTGTGATGCCATGAA 519
QY 703 AAGCATGCTGTAATGGCAGCAATAAATAGACAGACACACCCAGGCACTTCTGTTGT 762
Db 520 AAGGCTATGCTTCCCGGCAACAGCAGGTGATCATCACTCTAAGGACACACCCCTCATC 579
QY 763 TGTTCAGATTTGGAGTACCTTAAGATCTAGATCAAAATGTTTAAATTCAGAGGCGTT 822
Db 580 CACCAAAATTTGGAGGCTACTGAGATCTCAAAATCAAGTGTCTCCACGTGCCAGGCAAT 639
QY 823 TCAGATACCTTTTGCATATCTTGTATTAATTAAGATGAGGCTTCAAGGCTGCTAGAGTGC 882
Db 640 TCAGACACTTTTGCATATCTTGTATTAATTAAGATGAGGCTTCAAGGCTGCTAGAGTGC 699
QY 883 ACAAGGCACTTGGAGCAGTTTGTGAAGCGGAGCACTTGAATGAGGCACTTGAATGAGGCACT 942
Db 700 CAGCAAGCTTTGGAACAGTTGTTGAAGCGGAGCACTTGAATGAGGCACTTGAATGAGGCACT 759
QY 943 TSCAGCAAGTGTAAAGATGTTTCCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCT 1002
Db 760 TGTGGTGTGTTTCTCAGAGGCGCGGCTTCCAGAGCGTTTAACTTTACACCTCTGCC 819
QY 1003 AATGTTCTTACACTTCTCTGAAACGTTTTCGAAATTTTACCGGTGGAAATTTGCTAAG 1062
Db 820 AAGGTCCTCATCTTGTATTTGAGAGATTTCTCGATGTCAAGCAAGATTTGCAAG 879
QY 1063 GATGGAATATCCCTGAGTATCTTGTATTTCCGCGCATATATGCTCAACCCAGGAGAG 1122
Db 880 AATGTCAATATCTGAGTGCCTTGCATGACGACCATACATGTCTCAGACGAACACAGGA 939
QY 1123 CCAATGTCTAGCTTGTATGAGTGTGCTGCTGCTCCACACTGTTTAAATTTGCTGCTGCC 1182
Db 940 CTTCTGCTATGCTCTTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 999
QY 1183 CATTAATCTCTGCTACATAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1242
Db 1000 CATTAATCTCTGCTACATAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1059
QY 1243 GTATCTACAGTATATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1302
Db 1060 GTACCGGCTCTAGCATCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
QY 1303 AGGTCCCATGATGTGAATAATGGAGTGA 1331
Db 1120 CAGAAGAGTGAATGGAGAGACACAGTGA 1148

RESULT 14

US-10-371-905A-25
; Sequence 25, Application US/10371905A
; Publication No. US20030224969A1
; GENERAL INFORMATION:
; APPLICANT: Hahn, Chang

; APPLICANT: Liu, Hong
; TITLE OF INVENTION: Human Analogs of Murine Deubiquitinating Protease Genes
; FILE REFERENCE: US/2001/0185
; CURRENT APPLICATION NUMBER: US/10/371,905A
; CURRENT FILING DATE: 2003-02-20
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: human DUB8.1
US-10-371-905A-25

Query Match 17.8%; Score 430.2; DB 13; Length 1593;

Best Local Similarity 62.0%; Pred. No. 1.1e-127;

Matches 725; Conservative 0; Mismatches 423; Indels 21; Gaps 2;

QY 163 ATGGATGCAAGTTCTCCAGCTGGGCTGCTGCTCTTCAATGATGATGTCATCAATCAC 222
Db 1 ATGGGGGACACTCACTCTACTTGGAGGTGAGTGGCAGTTCACCACTTTTCAAAATC 60
QY 223 ACATTTTCTTTAGGACAGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 282
Db 61 ACATTTTCTCGGCAATATGCAAGCTTTTGTGAATTCCTCA---GCGGACTTCTCTCCCTGAG 117
QY 283 AATCAAAACCACTCACCAAAAGATCAAGCCTTAGGTGATGGCATTCGCTCTCCACAG 342
Db 118 AAGTCAACCACTCTCACTCTGAGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 177
QY 343 AAGTTCTTTTCCCACTCTGAGAGATTTGCTTAAAGTGGCAACAACTCATAGATTGA 402
Db 178 AGACGCTCGCTCCAGGAGAGCTTCTCTGAGTAGCAGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 237
QY 403 GCTGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCAATGCAAGCACTGCAAGTGTTTAAAC 462
Db 238 GCTGGGCTCCAGAAATTTGGGCAATACCTGCTAGGAGACGCTTCCCTGCTGCTGCTGCTGCTGCTG 297
QY 463 TACACACCACTTCTCCAAATTAATGCTATACATGAAACACTTCCAAACATGCTATGCA 522
Db 298 TACACACTGCTTCTCCAACTTACATGCTGCTCCGAGGAGCACTCTCAAAACATGCTAGCGT 357
QY 523 GAAGGCTTTTGTATGATGTGTAATGCAAGCACTATTTACCCAGGCACTTCACTATCCT 582
Db 358 CCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
QY 583 GGGGAGCTTTATTAACCAATGTTTGTATGATGAGATGCGGCTATAGCTAGGCACTTC 642
Db 418 GGCCA-----TGTCATCCAGCCCTCACAGGCACTTGGCTGCTGCTGCTTC 459
QY 643 CGTTTTTGGAAACCAATGAGATGCCATGAATTTCTTCAATATACACTGTTGATGCTATGCGAG 702
Db 460 CATAGAGGCAAGCAGTAAGATGCTCAATGAATTTCTCATGTTTCACTGCTGATGCCATGAA 519
QY 703 AAGCATGCTTGAATGGCAGCAATAAATAGACAGACACACCCAGGCACTTCTGTTGT 762
Db 520 AAGGCTATGCTTCCCGGCAACAGCAGGTGATGATCATCTGCAAGGACACACCCCTCATC 579
QY 763 TGTTCAGATTTTGGAGTACCTTAAGATCTAGATCAAAATGTTTAAATTTGCAAGGCGGTT 822
Db 580 CACCAAAATTTTGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639
QY 823 TCAGATACCTTTTGCATATCTTGTATTAATTAAGATGAGGCTTCAAGGCTGCTAGAGTGC 882
Db 640 TCAGACACTTTTGCATATCTTGTATTAATTAAGATGAGGCTTCAAGGCTGCTAGAGTGC 699
QY 883 ACAAGGCACTTGGAGCAGTTTGTGAAGCGGAGCACTTGAATGAGGCACTTGAATGAGGCACT 942
Db 700 CAGCAAGCTTTGGAACAGTTGTTGAAGCGGAGCACTTGAATGAGGCACTTGAATGAGGCACT 759
QY 943 TGCAGCAAGTGTAAAGATGTTTCCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCT 1002

Db 760 TGGCGTCTTTGTCCTCCAGAGGGCGCGCTCCACAGTTAACTTTACACACTTCTGCC 819
Qy 1003 AATGTTCTTACACTTTCTCTGAAAGCTTTTGCAGGTTTCCGCTGGGAAAAATGCTAAG 1062
Db 820 AAGGTCTCTCATCTTCTTGAAGAGATCTCTCGATGTCGAGGCAACAAACTTGCCAA 879
Qy 1063 GATGTGAATACCTTGATATCTTGATATTGGCCATATATGTCTCAACCCCAACGGAG 1122
Db 880 AATGTGAATATCTTGAGTGCTTGCATATGAGCCATACATGTCTCAGCAGAACACAG 939
Qy 1123 CCAATTGTCTAGCTTCTGATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1182
Db 940 CCTCTGTCTATGCTCTATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 999
Qy 1183 CATTACTTCTCTCATATAAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTG 1242
Db 940 CATTACTTCTCTCATATAAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTG 999
Qy 1183 CATTACTTCTCTCATATAAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTG 1242
Db 1000 CATTACTTCTCTCATATAAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTG 1059
Qy 1243 GTATCTACAGTATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1302
Db 1060 GTCACTGCTCTAGCATCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
Qy 1303 AGGTCCCATGATGTGAAATGGAGTGA 1331
Db 1120 CAGAAGAGTGAATGGAAAGACACAGTGA 1148

RESULT 15

US-09-372-348-10
; Sequence 10, Application US/09372348A
; Publication No. US20030028005A1
; GENERAL INFORMATION:
; APPLICANT: Hedrick, Joseph A.
; APPLICANT: Wang, Luquan
; APPLICANT: Zlotnik, Albert
; APPLICANT: Murgolo, Nicholas J.
; APPLICANT: Greene, Jonathan R.
; APPLICANT: Johnston, James A.
; APPLICANT: Bazan, J. Fernando
; TITLE OF INVENTION: Mammalian Proteins; Related Reagents and Methods
; FILE REFERENCE: DX0884K(2d)
; CURRENT APPLICATION NUMBER: US/09/372,348A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: primate
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1590)
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (837)
; OTHER INFORMATION: nucleotide 837 may be a or c; resulting amino acid
; OTHER INFORMATION: depends upon genetic code
US-09-372-348-10

Query Match
Best Local Similarity 62.0%; Pred. No. 1.2e-127; Length 1683;
Matches 725; Conservative 0; Mismatches 423; Indels 21; Gaps 2;
Qy 163 ATGGATGAGGTTCTGCCAGCTGGGGTGTGCTTCTTCAATGATGATGTCATCAATCAC 222
Db 1 ATGGAGGAGGACTCACTTACTTGGAGGTGAGTGGCAGTCAACCACTTTTCAAACTC 60
Qy 223 ACACCTTCTTAGACACAGTACTGGTGTGTGATTTATTCAGTTCATCTGTACTGAT 282
Db 61 ACATCTTCTCGGCCAGATGCACTTTTGTGTAATCCCA---CGGACTTCTCTCCCTGAG 117
Qy 283 AAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAG 342

Db 118 AAGTCAACACTCTCTATCTGAGGCGCGCTGTCACCTCTGTGATGATTTGCTCCTGTGCA 177
Qy 343 AAGTCTTTTTTCCCATCTGAGAGATTTTGTCTTAAGTGGCAACAAACTCATAGATTTGGA 402
Db 178 AGACAGCTTTGCTCCAGGGAGAGCTTCTCTGAGTAGCAGGAGACCTGCTGCGGTGGGG 237
Qy 403 GCTGGGCTCCAGATTTGGGCAATACCTGTTTGGCAATGACAGCACTGCTGCTGCTGCTG 462
Db 238 GCTGGGCTCCAGATATATGGGAAATACCTGCTACGAGAAAGCTTCCCTCAGTGGCTGCA 297
Qy 463 TACACACCACTCTTGGCAATTTACATGCTATACATGAACCACTCCAAACATGTTCATGCA 522
Db 298 TACACACCGCCCTTGGCAATTTACATGCTATGCTGCTCCGGGAGCACTCTCAAACTGTCAG 357
Qy 523 GAAGGCTTTTGTATGATGTGATCAATGCAAGCACATATTATCCAGGCACTCAGTAATCCT 582
Db 358 CCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 417
Qy 583 GGGAGAGTATTAACCAATTTGTGCTATCAATGAGATGCGCGGTATAGCTAGGACCTC 642
Db 418 GGTCA-----TGTCTATCCAGCCCTCACAGGCAATGGCTGCTGCTGCTGCTGCTGCT 459
Qy 643 CGTTTGGAAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCTGAG 702
Db 460 CATAGAGCAAGCAGGAGATGCCCATGATTTCTCATGTTCACTGTGTGATGCCATGAA 519
Qy 703 AAGCATGCTTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCACTTTGTT 762
Db 520 AAGCATGCTTTCCCGGCCACAAGCAGGTAGATCATCTCTTAAGGACACCACTCATC 579
Qy 763 TGTGAGATATTTGGAGGATACCTAAGATCTAGATCAATGTTTAAATGCAAGGCGCTT 822
Db 580 CACCAATATTTGGAGGCTGCTGGAGATCTCAATCAAGTGTCTCCACTGCGCACCGGAT 639
Qy 823 TCAGATATCTTTGATCCATATCTTGATATACTTGGAGATAAAGGCTGCTCAGAGTGTCT 882
Db 640 CAGACACTTTTGACCTTCTCTGGAATCGCCCTGGATATCCAGGAGCTCAGAGTGTCT 699
Qy 883 AACAGGCAATGGAGCAGCTTTGTGAAGCCGGAACAGCTTGTATGGAGAAACTCGTACAG 942
Db 700 AAGCAAGCTTTGGAACAGCTTGTGTAAGCCGGAAGAACTCAATGGAGAGATGCCATCAT 759
Qy 943 TGCAGCAAGTGTAAAGAGATGTTCCAGCTTCAAGAGTTCACCTATCCATAGATCTCTCT 1002
Db 760 TGTGCTCTTTGCTCTCAGAGGCGCGCCCTCCAAGACGTTAACTTTACACACTTCTGCC 819
Qy 1003 AATGTTCTTACACTTCTCTGAAAGCTTTTGCAAAATTTTACCGTGGAAAAATTCCTAAG 1062
Db 820 AAGTCTCTCATCTTCTGTTGTTGAAGAGATCTCCGATGTCAAGGCAACAACTTGCCAG 879
Qy 1063 GATGTGAATACCTTGAGTATCTTGAATATCGGCCATATATGTCTCAACCCCAACGGAG 1122
Db 880 AATGTGCAATATCTGAGTGCTTGCATGAGCCATACATGTCTCAGCAGAACACAGCA 939
Qy 1123 CCAATGCTAGCTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1182
Db 940 CTTCTTGTATGCTCTCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 999
Qy 1183 CATTACTTCTGCTACATAAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1242
Db 1000 CATTACTTCTCTTATGTCAAGCTCAGAAGCCAGTGTGATATAATGGATGATGCCAG 1059
Qy 1243 GTATCTACAGTATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1302
Db 1060 GTCACTGCTCTAGCATCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1119
Qy 1303 AGGTCCCATGATGTGAAATGGAGTGA 1331
Db 1120 CAGAAGAGTGAATGGAAAGACACAGTGA 1148

Search completed: August 17, 2004, 00:13:44

Job time : 1507 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:55:33 ; Search time 13 Seconds
(without alignments)
3104.182 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIVDKASESSPSAYQNQP.....AAESLEPPDAARSLFFPSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	4022	98.8	1198	1 UB42 HUMAN	Q9h9j4 homo sapien
2	1422.5	35.0	1121	1 UB36 HUMAN	O9p275 homo sapien
3	846.5	20.8	526	1 UB2W MOUSE	G61068 mus musculus
4	428.5	10.5	593	1 UB22 HUMAN	Q9u0t9 homo sapien
5	428	10.5	471	1 UB28 YEAST	F50102 saccharomyc
6	403	9.9	520	1 UB33 MOUSE	Q91w36 mus musculus
7	395.5	9.7	521	1 UB23 HUMAN	Q9y614 homo sapien
8	392	9.6	605	1 UB22 HUMAN	O75604 homo sapien
9	385.5	9.5	357	1 UB22 CHICK	O57429 gallus gall
10	383.5	9.4	353	1 UB22 MOUSE	O88623 mus musculus
11	358.5	8.8	449	1 UB28 SCHPO	Q09738 schizosacch
12	356.5	8.8	566	1 UB21 MOUSE	Q9qz16 mus musculus
13	352.5	8.7	366	1 UB46 HUMAN	O80v95 homo sapien
14	349	8.6	565	1 UB21 HUMAN	Q9uk80 homo sapien
15	343.5	8.4	792	1 UB2A YEAST	P53874 saccharomyc
16	342	8.4	355	1 UB12 HUMAN	O75317 homo sapien
17	336.5	8.3	712	1 UB44 HUMAN	Q9h0e7 homo sapien
18	332	8.2	805	1 UB25 YEAST	F39944 saccharomyc
19	324.5	8.0	1118	1 UB28 HUMAN	F40818 homo sapien
20	310	7.6	410	1 UB2X CAEBL	P34547 caenorhabdi
21	307.5	7.6	977	1 UB24 HUMAN	Q9u0u5 homo sapien
22	306	7.5	898	1 UB2E DROME	Q24574 drosophila
23	304	7.5	926	1 UB29 YEAST	F32571 saccharomyc
24	295.5	7.3	585	1 UB29 SCHPO	Q9p7v9 schizosacch
25	295.5	7.3	942	1 UB33 HUMAN	O8tev7 homo sapien
26	287.5	7.1	798	1 UB10 HUMAN	O14694 homo sapien
27	279	6.9	792	1 UB10 MOUSE	P52479 mus musculus
28	275.5	6.8	1142	1 UB40 HUMAN	Q9nves homo sapien
29	270.5	6.6	2547	1 FAFX HUMAN	Q93008 homo sapien
30	268.5	6.6	1140	1 UB40 MOUSE	Q8bw74 mus musculus
31	267	6.6	438	1 UB2A SCHPO	O60139 schizosacch
32	266.5	6.5	1102	1 UB27 HUMAN	Q93009 homo sapien
33	266.5	6.5	2555	1 FAFY_HUMAN	O00507 h probable

34	264.5	6.5	2559	1 FAFX_MOUSE	P70398 m probable
35	260	6.4	2778	1 FAF_DROME	P55824 drosophila
36	257	6.3	512	1 UB23 SCHPO	O94269 schizosacch
37	257	6.3	1129	1 UB2L SCHPO	Q9utt1 schizosacch
38	256	6.3	754	1 UB29 YEAST	P39967 saccharomyc
39	255.5	6.3	913	1 UB20_HUMAN	Q9y2k6 homo sapien
40	253	6.2	849	1 UB21 SCHPO	Q9usm5 schizosacch
41	251	6.2	1108	1 UB25 SCHPO	Q09879 schizosacch
42	250.5	6.2	1230	1 UB2F YEAST	P50101 saccharomyc
43	246.5	6.1	823	1 UB16 HUMAN	Q9y5t5 homo sapien
44	236.5	5.8	1042	1 UB38 HUMAN	Q8nb14 homo sapien
45	236.5	5.8	1042	1 UB38_MOUSE	Q8bw70 mus musculus

ALIGNMENTS

RESULT 1

UB42_HUMAN					
ID	UB42_HUMAN	STANDARD;	PRT;	1198 AA.	
AC	Q9H9J4;				
DT	10-OCT-2003 (Rel. 42, Created)				
DT	10-OCT-2003 (Rel. 42, Last sequence update)				
DE	Ubiquitin carboxyl-terminal hydrolase 42 (EC 3.1.2.15) (Ubiquitin thiolesterase 42) (Ubiquitin-specific processing protease 42)				
DE	(Deubiquitinating enzyme 42) (Fragment).				
GN	USP42.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,				
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,				
RA	Wagatsuna M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,				
RA	Takanashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,				
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,				
RA	Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y.,				
RA	Ninomiya K., Iwayanagi T.;				
RT	"NEDO human cDNA sequencing project."				
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.				
CC	-1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =				
CC	ubiquitin + a thiol.				
CC	-1- SIMILARITY: Belongs to peptidase family C19.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AK022759; BAB14232.1; -				
DR	MEROPS; C19.048; -				
DR	Genew; HGNC:20068; USP42.				
DR	InterPro; IPR001394; Peptidase_C19.				
DR	Pfam; PF00443; UCH_1.				
DR	PROSITE; PS00972; UCH_2_1; FALSE_NEG.				
DR	PROSITE; PS00973; UCH_2_2; 1.				
DR	PROSITE; PS0235; UCH_2_3; 1.				
KW	Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.				
FT	ACT_SITE 120 120 BY SIMILARITY.				
FT	ACT_SITE 362 362 BY SIMILARITY.				
FT	ACT_SITE 371 371 BY SIMILARITY.				
FT	NON_TER 1198 1198				
SQ	SEQUENCE 1198 AA; 130587 MW; CF0FDB5184AE9536 CRC64;				

Query Match 98.8%; Score 4022; DB 1; Length 1198;
Best Local Similarity 99.6%; Pred. No. 2.5e-209;
Matches 766; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 MTIVDKASESDPSAYQNPQSSSEAVSPGDMDSAGSAGVSSLNDVSNHTLSLGPVPGA 60
Db 1 MTIVDKASESDPSAYQNPQSSSEAVSPGDMDSAGSAGVSSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPQKQDQALGDIAPQKVLFPSEKICLKWOOTHRVAGLQNLGNTC 120
Db 61 VVYSSSVDPKSPQKQDQALGDIAPQKVLFPSEKICLKWOOTHRVAGLQNLGNTC 120
QY 121 FANAALQCLTTPPLANTMLSHEHSKTCBAGFCMCTMQAHITQALSNGPDGVKPMFVI 180
Db 121 FANAALQCLTTPPLANTMLSHEHSKTCBAGFCMCTMQAHITQALSNGPDGVKPMFVI 180
QY 181 NEMRIARHLRFGNQEDAHFLQYTVDMQKACLNGSNKLDHRHTQATTLVQIFGGLYLS 240
Db 181 NEMRIARHLRFGNQEDAHFLQYTVDMQKACLNGSNKLDHRHTQATTLVQIFGGLYLS 240
QY 241 RVKCLNCKGVSDTTPPYLDITLEIKAAQSVNKALBQFVKPQLDGENSYKSKCKKMPVA 300
Db 241 RVKCLNCKGVSDTTPPYLDITLEIKAAQSVNKALBQFVKPQLDGENSYKSKCKKMPVA 300
QY 301 SKRFTIHRSSNVLTSLKRFANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
Db 301 SKRFTIHRSSNVLTSLKRFANFTGGKIAKDVKYPEYLDIRPYMSQNGEPIVYVLYAVL 360
QY 361 VHTGFNCAGHYFCYIKASNGLIWQNDISIYSTDIRSVLSQAYVLYIRSHDVKNKGE 420
Db 361 VHTGFNCAGHYFCYIKASNGLIWQNDISIYSTDIRSVLSQAYVLYIRSHDVKNKGE 420
QY 421 LTHPTHSPGQSPRPIVSQVVTNKAAPGFTGPOLPSHMIKNPPLHNGTGPKLDTSPSS 480
Db 421 LTHPTHSPGQSPRPIVSQVVTNKAAPGFTGPOLPSHMIKNPPLHNGTGPKLDTSPSS 480
QY 481 MSSPNGSSVNRAGPVNASVQNMVNRSSVIPBHPKKQKITISIHNLKPVRCQSQPN 540
Db 481 MSSPNGSSVNRAGPVNASVQNMVNRSSVIPBHPKKQKITISIHNLKPVRCQSQPN 540
QY 541 LHSNSLENTKVPVSTTNSAVQSTNSASTVSSKTKPLPRSESCSQPMVNGSKLIN 600
Db 541 LHSNSLENTKVPVSTTNSAVQSTNSASTVSSKTKPLPRSESCSQPMVNGSKLIN 600
QY 601 SSVLPYGAESSEDSDESKGLGKENGIGTIVSSHSPQDADEATPHEIQEPTTLNGA 660
Db 601 SSVLPYGAESSEDSDESKGLGKENGIGTIVSSHSPQDADEATPHEIQEPTTLNGA 660
QY 661 NSADSDPKENGLAPDAGSCQGPALHSENPFKANGLPGLKMPAPLLSLPEDKILETF 720
Db 661 NSADSDPKENGLAPDAGSCQGPALHSENPFKANGLPGLKMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKGTSDMSAPGERGPPEDRDAEPQGSFAESLEPDAASL 769
Db 721 RLSNKLKGTSDMSAPGERGPPEDRDAEPQGSFAESLEPDAASL 769

RESULT 2
UB36 HUMAN
ID UB36 HUMAN STANDARD; PRT; 1121 AA.
AC Q9P275; Q9NDM8; Q9NVC8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 36 (EC 3.1.2.15) (Ubiquitin
thiolesterase 36) (Ubiquitin-specific processing protease 36)
DE (Deubiquitinating enzyme 36).
GN USP36 OR KIAA1453.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20277482; PubMed=10819331;
```

```
RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XVII.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro.";
RL DNA Res. 7:143-150(2000).
RN [2]
RP SEQUENCE OF 1-954 FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Negai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 494-1121 FROM N.A.
RC TISSUE=Testis;
RA Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SUBCELLULAR LOCATION
RX MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus.";
RL Mol. Biol. Cell 13:4100-4109(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
ubiquitin + a thiol.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB040886; BAA95977.1; ALT_INIT.
DR EMBL; AK001671; BAA91823.1; -.
DR EMBL; AL838835; CAD38693.1; -.
DR SWISS-2DPAGE; Q9P275; HUMAN.
DR Genbank; HGNC:20062; USP35.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2.1; 1.
DR PROSITE; PS00973; UCH_2.2; 1.
DR PROSITE; PS0235; UCH_2.3; 1.
DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family;
KW Nuclear protein.
FT ACT_SITE 131 BY SIMILARITY.
FT ACT_SITE 373 BY SIMILARITY.
FT ACT_SITE 382 BY SIMILARITY.
FT CONFLICT 573 D -> G (IN REF. 2).
FT CONFLICT 806 R -> Q (IN REF. 3).
FT CONFLICT 828 C -> R (IN REF. 3).
FT CONFLICT 958 K -> KKK (IN REF. 3).
SQ SEQUENCE 1121 AA; 122626 MW; 2476F5128CBAS5 CRC64;

Query Match 35.0%; Score 1422.5; DB 1; Length 1121;
Best Local Similarity 41.1%; Pred. No. 2.3e-69;
Matches 335; Conservative 110; Mismatches 244; Indels 127; Gaps 25;

QY 1 MTIVDKASESDPSAYQNPQSSSEAVSPGDMDSAGSAGVSSLNDV----- 47
Db 1 MPVLDKLEAL-----KPGKSDADDGELGKLASSAKVKVLLKIEFEPASKFSYQL 53
QY 48 ----SNHTLSLGPVPGAVVY-SSSVDPKSPQKQDQALGDIAPQKVLFPSEKICLK 102
Db 54 EALKSKYVLLNPKTEGASRHKSGDDPPARRQSGSEHTYESCGDGPAPQKVLFPTELSLR 113
QY 103 WOOTHRVGAGLONLGNTPFANALQCLTYTPPLANTMLSHEHSKTCBAGFCMCTMQAH 162
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db 114 WERVPRVAGLHNLGNTCFNLATIQCLTTPPLANVLLSKEHARSCHQSGFCMLCWMQNH 173

QY 163 ITOALSNPGDVIKPMFVINEMRIARHLRFGNQEADAEFLQYIVDAMQKACLSNKLDR 222

Db 174 IVOAFANSNAIKPVSFIRDLKIAHFRFGNQEADAEFLRYTIDAMQKACLSNCAKLR 233

QY 223 HTQATTLVQVIFGGYLSRSVKLNCCKGVSDTDPYLDITLEIKAAQSVNKALQFVKPEQ 282

Db 234 QTOATTLVHQVIFGGYLSRSVKCSVKCSVSDTDPYLDVALEIRQANIVRALEFLKADV 293

QY 283 LQGENSKCKCKKMPVAPSKRFTIHRSSNNVLTSLKRFANFTGGKIAKOVKYPEYLDIRP 342

Db 294 LSGENAYMAKCKKVPKASKRFTIHRSSNNVLTSLKRFANFTGGKITKDVGYEFLNIRP 353

QY 343 YNSQNGEPTVYLYAVLHGTGNCHAGHYFCVYKASGLWYQNDSTVSTDIRSVLSQ 402

Db 354 YMSQNGDPMVYGLYAVLVHSGYCVYKASGLWYQNDSTVSTDIRSVLSQ 413

QY 403 QAYVLYFIRSHDVKNKGELTHPETHSPG-OSSPRPVISQSVVTKQAAPGFIGPO-LPSHM 460

Db 414 QAYVLYFIRSHDVKNKGELTHPETHSPG-OSSPRPVISQSVVTKQAAPGFIGPO-LPSHM 460

QY 461 IKNPPLHNGTGPL--KDTFSSMSPNNGSNVNRASPVNASASVQWNSVNRSSVIEHPK 518

Db 454 KKNIGNGLIISPLTKRQKQSDGTMKKPHTEEI--GVPIRNGSTLGLKQNGCIPPKLPS 511

QY 519 KQKITIISHNKLPVRCQSQPNLHNSLENP-----TKPVPS-----SITINSVOSTNAST 571

Db 512 GS-----PSPKLSQTPHTMPTILDDPGKKVPAPQPHSPRTAQGLFGTSNSN- 560

QY 572 MSVSGKVTXKPIPRSESCSPVNMNGKLSNVLV-PYGAESSESDSEKGLKENGIGT 630

Db 561 ---SSRSQORQGSWDSRDVLSTSPKLLATATANGHLK-----GNDSAGLDR-RGSS 612

QY 631 IVSSHSPQDADEERATPHELQEPMTLNGANSADSDPKENGLAPDGASCOQPALHSE 690

Db 613 SSPEHSASSDSTKAPOTPRS-----GAHLCDSE-----TNC-STAGHSK 652

QY 691 NPPAKANGLPGLM-----PAPLSLEDKILEFRLSNKLKSGTDEMSAPGAER 740

Db 653 TPSPGADSKTVKLSVLSNTTTEPASTWSPPAK-----KLALSAKASTLWRATG--- 704

QY 741 GPPEDRADBPQGSAAESLEEP-----DAAASLFP 771

Db 705 -----NDLRPPPPSSD-LTHPMKTSHPVASTWP 734

RESULT 3

UBPW MOUSE STANDARD; PRT; 526 AA.

AC Q61058;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Ubiquitin carboxyl-terminal hydrolase DUB-1 (EC 3.1.2.15) (Ubiquitin thiolesterase DUB-1) (Ubiquitin-specific processing protease DUB-1)

DE (Deubiquitinating enzyme 1).

GN DUB1 OR DUB-1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN

RP SEQUENCE FROM N.A., AND MUTAGENESIS OF CVS-60.

RA MEDLINE=96194957; PubMed=8622927;

RX Zhu Y., Carroll M., Papa F.R., Hochstrasser M., D'Andrea A.D.;

RT "DUB-1, a deubiquitinating enzyme with growth-suppressing activity.";

RL Proc. Natl. Acad. Sci. U.S.A. 93:3275-3279(1996).

CC -!- FUNCTION: Has growth-suppressing activity, induces arrest in G1 phase upon controlled expression.

CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O = ubiquitin + a thiol.

CC -!- INDUCTION: By interleukin-3.

CC -!- SIMILARITY: Belongs to peptidase family C19.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; U41636; AAC52532.1; -.

DR PIR; JC6133; JC6133.

DR MEROPS; C19.031; Dub1.

DR MGD; MGI:107699; Dub1.

DR GO; GO:0004843; Fubiquitin-specific protease activity; IDA.

DR GO; GO:0016579; P:protein deubiquitination; IDA.

DR InterPro; IPR001394; Peptidase_C19.

DR Pfam; PF00443; UCH; 1.

DR PROSITE; PS00972; UCH_2_1; 1.

DR PROSITE; PS00973; UCH_2_2; 1.

DR PROSITE; PS02335; UCH_2_3; 1.

KW Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.

FT ACT SITE 60 60 BY SIMILARITY.

FT ACT SITE 298 298 BY SIMILARITY.

FT ACT SITE 307 307 BY SIMILARITY.

FT MUTAGEN 60 60 C->S: LOSS OF ACTIVITY.

SQ SEQUENCE 526 AA; 59073 MW; 263AA7B7579694EA CRC64;

Query Match 20.8%; Score 846.5; DB 1; Length 526;

Best Local Similarity 37.1%; Pred. No. 9.6e-39;

Matches 186; Conservative 89; Mismatches 152; Indels 75; Gaps 12;

QY 76 PKDQALGDIAPP-----QKVLFPSEKICLKWQOQTHRVGAGLQNTGFANAAQ 127

Db 8 PEADPALSPDAPLHQDEAQQVVEELTVNGKSHLSWESQPGCGQLQNTGNSCYLNAALQ 67

QY 128 CLTYTPPLANYMLSEHSKTHAEGFCMCTMQAHTQAL--SNPGDVIKPMFVINEMR 185

Db 68 CLTHTPPLADYMLSQHSQTCSPGCKLCAMEALVTQSLHSHSGDVNMPKSHILTS--- 124

QY 186 IARHLRFGNQEDAEFLQYIVDAMQKACLSNKLDRHTQATTLVQVIFGGYLSRSVKCL 245

Db 125 -AFHKH--QEDAEHFLMFTLETMHESCLQVHRQKPSSEDSPIHDFGGWRSQIKCL 181

QY 246 NCKGVSDTDPYLDITLEIKAAQSVNKALQFVKPEQLDGENSYKSKCKKMPVAPSKRFT 305

Db 182 LCQGTSDTYDRELDIPLDISSAQSVKQALWDTEKSEELCGDNAYYCGKCKQKMPASKTLH 241

QY 306 IHRSSNVLTSLKRFANFTGGKIAKOVKYPEYLDIRPYMSQNGEPTVYLYAVLVHTGF 365

Db 242 VHIAPKVLMVNLNRFSAFTGNKLDKRVSYPEFLDLKPYLSEPTGGPLPYALYAVLVHGA 301

QY 365 NCHAGHYFCYIKASNGLWYQNDSTVSTDIRSVLSQQAAYLVFYR-----SHDVKN 418

Db 302 TSHSGHYCCVKGAGKWKMDTKVRCDSVLNENAYLVFYVQANLKQVSDIMPEG 361

QY 419 --GELTHPTH-----SPGSSPRPVISQ-----RVVTNQ 446

Db 362 RINEVLDPEYQLKGRKKKKKSPFTDLGEPFCENRDKRAIKETSLGKGLVQEVNHHK 421

QY 447 AAPGF-----IGPOLPSHMLKPPHNLGTGKPLKDTFSSS--MSSPNGNSVNRASP 495

Db 422 AGQKHGNTKLMPPQKHQ--KAGQNLNRTNEVELDPAIVIHQPRSTANWGRSDSPDKENQ 480

QY 496 -----VNASAYQNSVNR 509

Db 481 PLHNADRLITSQGVNTWQLCR 502

RESULT 4

UB22 HUMAN STANDARD; PRT; 593 AA.

ID UB22 HUMAN

AC Q9UFT9;

DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 22 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 22) (Ubiquitin-specific processing protease 22)
 DE (Deubiquitinating enzyme 22) (Fragment).
 GN USP22 OR KIAA1063.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AB028986; BAA83015.1; -
 CC MEROPS; C19.035; -
 CC Genew; HGNC:12621; USP22.
 DR InterPro; IPR001394; Peptidase_C19.
 DR InterPro; IPR001607; Znf_UBP.
 DR Pfam; PF00443; UCH; 1.
 DR Pfam; PF02148; Zf-UBP; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS02035; UCH_2_3; 1.
 KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT SITE 253 BY SIMILARITY.
 FT ACT SITE 539 BY SIMILARITY.
 FT ACT SITE 547 BY SIMILARITY.
 FT ACT SITE 547 BY SIMILARITY.
 SQ SEQUENCE 593 AA; 66605 MW; 862ED0E3ED746BEF CRC64;
 Query Match 10.5%; Score 428.5; DB 1; Length 593;
 Best Local Similarity 30.4%; Pred. No. 3.6e-16;
 Matches 104; Conservative 55; Mismatches 138; Indels 45; Gaps 6;
 QY 112 GLQLNGNTCFANAALQCLTYTPPLANYMLSHSKTC--HAEGFCMMCTMQ-- 171
 DB 245 GLINLNGTCFNCVIALQTHPLDLDFLSRHRCEMSPSSCLVCMSSLFQEPYSGRH 304
 QY 172 DVIKPFMFINEMRRIARHLRFGNQDEAHEFLQYTVDMQKACLNGSN--KLDRLHTQATLL 229
 DB 305 SPHIPYKLLHLVWTHARHLAGVEQQAHEFLIALDLVLRHCKGDDNGKANNPNHNCI 364
 QY 230 VCQIFGGYLRVRKCLNCKGVSDFDPPYDITL-----I 264
 DB 365 IDQLFTGLQSDVTCQVCHGVSTTIDPFDWISLDLFGSSTFFWPLSPSGEGNVNGSHV 424
 QY 265 KAAGSVNKALEOFVKPEQLDGENSVKCSKCKKMWVPASKRFTTHRSSNVLTLSLKEFANFT 324
 DB 425 SGTTLTDLCLRRFTPEHLGSSAKIKCSGCHSYQESTQLITMKLPIVACFLKFEHSA 484
 QY 325 --GGKIAKDVKYPEYLDIRPYM-----SQPNGE-----PIYVYLYAVLHTGNC 367
 DB 485 KLRKITYYVFPFLDTPFWASSKSRMNGQYQQTDSLNDNDNKYSLEAVVNHQG--TL 543

QY 368 HAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQQAIVLFY 409
 DB 544 ESGHYTFIRQHKDQWPKCDALITKASIKDVLDSGLLYLFY 585
 RESULT 5
 UBP8_YEAST STANDARD; PRT; 471 AA.
 ID UBP8_YEAST
 AC P50102;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 8) (Ubiquitin-specific processing protease 8)
 DE (Deubiquitinating enzyme 8).
 GN UBP8 OR YMR223W OR YMR959.05.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313268; PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jagels K., Iye G., Moute S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT XIII.";
 RL Nature 387:90-93(1997).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z49939; GAA90194.; -
 CC FIR; S57591; S57591.
 CC GenOnline; 142898; -
 CC MEROPS; C19.0PW; -
 CC SGO; S0004836; UBP8.
 DR GO; GO:0000124; C:SAGA complex; IDA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR InterPro; IPR001607; Znf_UBP.
 DR Pfam; PF00443; UCH; 1.
 DR Pfam; PF02148; Zf-UBP; 1.
 DR SMART; SMO0290; Znf_UBP; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS02035; UCH_2_3; 1.
 KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT SITE 146 146 BY SIMILARITY.
 FT ACT SITE 419 419 BY SIMILARITY.
 FT ACT SITE 427 427 BY SIMILARITY.
 SQ SEQUENCE 471 AA; 53633 MW; BC632F12FBD0F73C CRC64;
 Query Match 10.5%; Score 428; DB 1; Length 471;
 Best Local Similarity 31.4%; Pred. No. 2.9e-16;
 Matches 104; Conservative 68; Mismatches 125; Indels 34; Gaps 11;
 QY 111 AGLNLGNTCFANAALQCLTYTPPLANYMLSHSKTC--HAEGFCMMCTMQ-- 160
 DB 137 SGLNMGSTCFMSSILQCLIHNPYFIRHSMQIHSNCKVSRSPDKFCSCALDKIVHLYG 196
 QY 161 -AHITQALSNPGDVIKIP--MFVINEMRRIARHLRFGNQDEAHEFLQYTVDMQKAC-- 213

Db 197 ALNTKQASSTSTNRQTQFIYLLTCWKINQNLQAGYSQDRAHEFWQFIINQIHQSVILD 256
 QY 214 LNSGNKLDLH-HQATTLVCQIFGGYLRVSRVCLNCKGVS-DTFDPYLDITLLEIKAAQSV 270
 Db 257 LPNAKEVSRANNKQCEIVHTVFEGLSSIVCPGQNNKSTTIDPFLDLSLDIKDKKXL 316
 QY 271 NKALQFVKPEOLDGNSYKCKKQWVASKFTTHRSNVLTLSKRFANFTGG---K 327
 Db 317 YECLOSFHKEQKQDFN-YHCGECNSTQDAIKQLGHKLPLSVLQLKRFPHLLNGSNRK 375
 QY 328 IAKDVKKPYLDLRPMS-----OPNGE--PIVVVLVAVLVHTGFNCHAGHYFCYIKA 378
 Db 376 LDDFIEFFPYLNMKNVCSYTKDKHSENGKVPDIIVELIGIVSHKG-TYNEGHYIAFCKI 434
 QY 379 SGNLWQMNDSIVSTSDIRSVLSQQAYVLFY 409
 Db 435 SGGQWFKFNDMSVSSISQBEVLKEQAYLLFY 465

RESULT 6

UBP3 MOUSE STANDARD; PRT; 520 AA.

AC Q91W36;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin
 thiolesterase 3) (Ubiquitin-specific processing protease 3)
 DE (Deubiquitinating enzyme 3).
 GN USP3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Forelimb;
 RX MEDLINE=22354683; PubMed=12466851;
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
 RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
 RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
 RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
 RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
 RA Birney E., Hayashizaki Y.;
 RA "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573 (2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins P.S., Wagner L., Shemmen C.M., Schuler G.D.,
 RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -!- SIMILARITY: Belongs to peptidase family C19.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AK031143; BAC27276.1; -;
 DR EMBL; BC017156; AAH17156.1; -;
 DR MEROPS; C19.026; -;
 DR MGD; MGI:2152450; Usp3.
 DR InterPro; IPR001394; Peptidase_C19.
 DR InterPro; IPR001607; Znf_UBP.
 DR Pfam; PF00443; UCH; 1.
 DR Pfam; PF02148; zf-UBP; 1.
 DR PROSITE; PS00972; UCH 2_1; 1.
 DR PROSITE; PS00973; UCH 2_2; 1.
 DR PROSITE; PS0235; UCH 2_3; 1.
 KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT SITE 168 168 BY SIMILARITY.
 FT ACT SITE 462 462 BY SIMILARITY.
 FT ACT SITE 471 471 BY SIMILARITY.
 SQ SEQUENCE 520 AA; 58868 MW; E160B48ECA47483D CRC64;
 Query Match 9.9%; Score 403; DB 1; Length 520;
 Best Local Similarity 29.9%; Pred. No. 7.3e-15;
 Matches 115; Conservative 51; Mismatches 117; Indels 102; Gaps 13;
 QY 112 GLQNLGTCFANAALQCLT-----YTPPL-----ANYMIS 141
 Db 160 GLRLNLGTCFNAILQSLNIFQCYFKELPAVELRNGKTRAGTYHTRSGDSNVSLV 219
 QY 142 HEHSKTCHAEGCMCTQQAHTQALSNPGDVIKPMFVINEMERIARHLRFQNGDAHEF 201
 Db 220 EEFKRT-----LCALWQGSQTAFS-----PESLFYVWVKIMPFRGYQQDAHEF 264
 QY 202 LOY-----TVDMOKACLNGSKNKLDRHTQATLYVCOIFG 235
 Db 265 MYRLDLHLHLELQGGFNGVSRSAIQENSTLSASNKCCING-----ASTVTAIFG 315
 QY 236 GYLRSRVKCLNCKGVSDFDPYLDITLII-----KAAQSVNK-----ALEQVKEP 281
 Db 316 GLQNEVNCCLIGTSEKRFDPFLDLSLQIPQFSKRSKNQNGPVCSLRDCIRSFDTLE 375
 QY 282 QLDGNSYKCKKQWVASKRFTTHRSNVLTLSLXRF--ANFTGGIAKDVKYP-EYL 338
 Db 376 ELDETELYMCHCKKKQSKTKFQKLPKALCLHLKRFHWTAYLRNKNVDVTVQFLRGL 435
 QY 339 DIRPYMSQPNG---EPTVVVLVAVLVHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSD 395

```

Db 436 DMKYLLEPNSGDSCLYDLAAVVHGGVSGSHYAT-AVHEGRWFHENDSTVITVD 494
QY 396 IRSVLSQOAYVLFYIRSHDVKNGE 420
Db 495 EETVGRKAKAVILFYV-BRQARAGAE 518

RESULT 7
UBP3 HUMAN
ID _UBP3 HUMAN STANDARD; PRT; 521 AA.
AC Q9Y6I4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 3 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 3) (Ubiquitin-specific processing protease 3)
DE (Deubiquitinating enzyme 3).
GN USP3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99410422; PubMed=10480896;
RA Sloper-Mould K.E., Eyre H.J., Wang X.-W., Sutherland G.R.,
RA Baker R.T.;
RT "Characterization and chromosomal localization of USP3, a novel human
RT ubiquitin-specific protease."
RL J. Biol. Chem. 274:26878-26894 (1999).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- TISSUE SPECIFICITY: Expressed in all tissues examined, with
CC strongest expression in pancreas.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF073344; AAD42992.1; -.
CC MEROPS; C19.026; -.
CC Genew; HGNC:12626; USP3.
CC MIM; 604728; -.
CC GO; GO:0004843; F:ubiquitin-specific protease activity; TAS.
CC InterPro; IPR001394; Peptidase_C19.
CC InterPro; IPR001607; Znf_UBP.
CC Pfam; PF00443; UCH; 1.
CC Pfam; PF02148; zf-UBP; 1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS0235; UCH_2_3; 1.
CC Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
KW ACT SITE 168 BY SIMILARITY.
FT ACT SITE 463 BY SIMILARITY.
FT ACT SITE 472 BY SIMILARITY.
SQ SEQUENCE 521 AA; 59097 MW; EF5CF39BA5482D67 CRC64;

Query Match 9.7%; Score 395.5; DB 1; Length 521;
Best Local Similarity 30.3%; Pred. No. 1.8e-14;
Matches 117; Conservative 48; Mismatches 118; Indels 103; Gaps 14;

QY 112 GLQNLGNTCFNANALQCLIT-----YTPPLA-----NYWLS 141
Db 160 GLRLNGNTCFNALLQSLSNIEQCCYFKELPAVELNGKTAGRTVHTRSGDNNVSLV 219
QY 142 HEHSEKTHAEFGCMWCTMQAHIQTALSNPGVDIKPMFVFNMRRIARHLRFGNEDAHEF 201

```

```

Db 220 EEFRKT-----ICALWQGSOTAFS-----PESLFYVWVKIMFNGFRGQQOADAHEF 264
QY 202 -----LOYTVDAQ-----KACLNGSNKLDHRHTQATTLVCOIF 234
Db 265 NALPFGPPTLGNFRAVSTVFPAQOFCRRILLCLQVKNCCING-----ASTVVTAF 315
QY 235 GGYLRSRVKCLNCKGVSDFDPYLDITLFI-----KAAQSVNK-----ALEQFVKP 280
Db 316 GGIQNEVNCILIGTESRKDFDPLDISLIPQFSRKRKNQENGPVCSLDCRLSRSTD 375
QY 281 EQLDGENSYKSKCKKMPASKRFTTHRSSNVLTLSLKR-ANFTGGKIADKVYK- 337
Db 376 EELDELYMCHCKKKQKSTKFKWIKLQKLVCLHLKRFHTAYLRKNKVDVYFEFLRG 435
QY 338 LDRPYMSQPNQ---EPVVVLYAVLVHTGFCNCHAGHYFCYIKASNLGLWQMNDISVTS 394
Db 436 LDMKWYLLEPNSGDSCLYDLAAVVHGGVSGSHYAT-AHEGRWFHENDSTVITVD 494
QY 395 DIRSVLSQOAYVLFYIRSHDVKNGE 420
Db 495 DEETVGRKAKAVILFYV-EHQAKAGSD 518

RESULT 8
UBP2 HUMAN
ID _UBP2 HUMAN STANDARD; PRT; 605 AA.
AC Q756Q4; Q9B021;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
GN USP2 OR UBP41.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RA Gong L., Yeh E.T.H.;
RT "Cloning and expression of the human and mouse UBP41."
RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lymph;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg E., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.E., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.N., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;

```



```
CC CC      Isoid=O75604-1; Sequence=Displayed;
CC CC      Name=2;
CC CC      Isoid=O75604-2; Sequence=VSP_005256, VSP_005257;
CC CC      Note=No experimental confirmation available;
CC CC      -!- SIMILARITY: Belongs to peptidase family C19.
CC CC      -----
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC      -----
CC CC      EMBL: AF079564; AAC28392.1; -
CC CC      EMBL: BC002854; AAH02854.1; -
CC CC      EMBL: BC002955; AAH02955.1; -
CC CC      MEROPS: C19.013; -
CC CC      Genew: HGNC:12618; USP2.
CC CC      MIM: 604725; -
CC CC      GO: GO:0004197; F:cysteine-type endopeptidase activity; TAS.
CC CC      InterPro: IPR001394; Peptidase_C19.
CC CC      Pfam: PF00443; UCH; 1.
CC CC      PROSITE: PS00972; UCH_2_1; 1.
CC CC      PROSITE: PS00973; UCH_2_2; 1.
CC CC      PROSITE: PS50235; UCH_2_3; 1.
CC CC      KW      Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family;
CC CC      Alternative splicing.
CC CC      FT      ACT_SITE 24 24      BY SIMILARITY.
CC CC      FT      ACT_SITE 297 297      BY SIMILARITY.
CC CC      FT      ACT_SITE 305 305      BY SIMILARITY.
CC CC      FT      VARSPLIC 1 252      Missing (in isoform 2).
CC CC      FT      VARSPLIC 253 258      /FTID:VSP_005256.
CC CC      FT      CONFLICT 594 594      L -> H (IN REF. 1).
CC CC      FT      CONFLICT 602 605      PSRM -> TSPI (IN REF. 1).
CC CC      SQ      SEQUENCE 605 AA; 68071 MW; AFF4DA9344D21812 CRC64;
CC CC      -----
CC CC      Query Match          9.6%; Score 392; DB 1; Length 605;
CC CC      Best Local Similarity 27.5%; Pred. No. 3.4e-14;
CC CC      Matches 123; Conservative 63; Mismatches 162; Indels 100; Gaps 16;
CC CC      -----
CC CC      6 KASESDPSAYQNPQS-SEAVSPGMDAGSAGVSLINDVSNLTL---SLGPVPGAV 61
CC CC      205 KGSASQVPS-QAPPSRVPEITSP-----TYRPIGRVTLWETGKGQAPGP- 247
CC CC      62 VYSSSVDPKSPQKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCF 121
CC CC      248 --SRSSSPGRDGMNKSQAQL-----AGLNLGNTCF 277
CC CC      122 ANAALQCLTYTPPLANYMLSHESKTCCH--ABGFCMCTMQAHITQAL---SNPGDVIKPM 177
CC CC      278 MNSILQCLNTRELRYCLQRLYMDLHGHSNAHTALVEEFAKLIQTIWTSPPNDVSPS 337
CC CC      178 FVINEMRIARHLRFNGNDAHEFLQYTVDMQKACLNGSKLDRHT-----224
CC CC      338 EFKTIQRYAPRFVGYNQDAQEFRLFDGLH-----NEVNRVLRPKNPENLDHL 390
CC CC      225 -----QATLVCOIFGGYLRSRVKCLNCKGVSDTFDPYLDITILEIK--- 265
CC CC      391 PDDEKGRQWRKYLREDSRIGDLFVGQLKSLTCTDCGYCVTFDPFDWLSLPTAKRY 450
CC CC      266 AAQSVNKALEQFVKPEOLDGENSYKSKCKMVPASKRFTIHRSSNVLTLSLKRFP--ANF 323
CC CC      451 PEVTLMDCMLFTKEDVLDGDEKPTCCRCRGRKRCIKFESIQRFPKILVLHKKFSESRI 510
CC CC      324 TCGKIADKVP- EYLDIRPYMSQNGEPIVVVLAVLVHTGPNCHAGFYCYIKA-SNG 381
CC CC      511 RTSKLTTFVNFPLRLDLREFASE-NTNHAVYNLVAVSNHSG-TTMGGHYTAICRSPGTG 568
CC CC      382 LWYQMNDSIVTSDIRSVLSSQQAIVLFY 409
CC CC      -----
Db      569 EWHTFNDSSVTPMSSSQVRTSDAYLLFY 596
RESULT 9
UBP2_CHICK
ID_UBP2_CHICK STANDARD; PRT; 357 AA.
AC OS7429;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
DE (deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
GN USP2 OR UBP41.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=97467343; PubMed=9325273;
RA Baek S., Choi K.S., Yoo Y.J., Cho J.M., Baker R.T., Tanaka K.,
RA Chung C.H.;
RT "Molecular cloning of a novel ubiquitin-specific protease, UBP41, with
RT isopeptidase activity in chick skeletal muscle.";
RL J. Biol. Chem. 272:25560-25565(1997).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF016107; AAC13729.1; -
CC MEROPS: C19.013; -
CC InterPro: IPR001394; Peptidase_C19.
CC Pfam: PF00443; UCH; 1.
CC PROSITE: PS00972; UCH_2_1; 1.
CC PROSITE: PS00973; UCH_2_2; 1.
CC PROSITE: PS50235; UCH_2_3; 1.
CC KW      Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
CC FT      ACT_SITE 28 28      BY SIMILARITY.
CC FT      ACT_SITE 301 301      BY SIMILARITY.
CC FT      ACT_SITE 309 309      BY SIMILARITY.
CC SQ      SEQUENCE 357 AA; 40931 MW; EC39E6454937C455 CRC64;
CC -----
Query Match          9.5%; Score 385.5; DB 1; Length 357;
Best Local Similarity 30.8%; Pred. No. 4e-14;
Matches 109; Conservative 52; Mismatches 134; Indels 59; Gaps 12;
-----
QY 112 GLQNLGNTCFANALQCLTYTPPLANYMLSHESKTCCHAEFGCMCTMQ--AHITQAL-- 167
DB 20 GLRLNLGNTCFMNSILQCLNTRELRYCLQRLYMDLNNNSRMTALMGEFAKLQLLWT 79
QY 168 SNPGDVIKPMFVINEMRIARHLRFNGNDAHEFLQYTVDMQKACLNGSKLDR----- 222
DB 80 SSPNDVSPSEFKTIQRYAPRFVGYNQDAQEFRLFDGLH-----GEVNRVLRP 132
QY 223 HQQATTL-----VCQIFGGYLRSRVKCLNCKGVSDTFDPYLD 259
DB 133 RANADTLDLHPDEKSRQWRKYLREDSRIGDLFVGQLKSLTCTDCGYCVTFDPFDW 192
QY 260 ITLEIK-----AAQSVNKALEQFVKPEOLDGENSYKSKCKMVPASKRFTIHRSSNV 315
DB 193 LSLPLPKGYGEVTLMDCLFLTKEDVLDGDEKPTCCRCRGRKRCIKFESIQRFPKILVL 252
```

```

QY 316 SLKRF--ANFTGGKIAKDVYP-EYLDIRPYMQNGEPIVYVLYAVLVHTGFNCAGHY 372
Db 253 HLKRFSEARIRASKLTTFNFPFLKDLDRFASQ-SCNHAAYNLVAVSNHSG-TTMGGHY 310
QY 373 FCYIKAS-NGLWYQMNDSIVSTDIRSVLSQAAVLFYIRSHDVKNKGELTHPT 425
Db 311 TAYCKPISSEHNSFNDVRTPMSSSHVRSDDAYLLFY-----ELASPS 354

RESULT 10
UBP2_MOUSE
ID UB2_MOUSE STANDARD; PRT; 353 AA.
AC O88623;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
DE thioesterase 2) (Ubiquitin-specific processing protease 2)
DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
GN USP2 OR UBPA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Gong L., Yeh E.T.H.;
RT "Cloning and expression of the human and mouse UBPA1.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF079565; AAC28393.1; --
CC MEROPS; C19.013; --
CC MGD; MGI:1858178; Usp2.
CC InterPro; IPR001394; Peptidase_C19.
CC Pfam; PF00443; UCH_1.
CC PROSITE; PS00972; UCH_2_1; 1.
CC PROSITE; PS00973; UCH_2_2; 1.
CC PROSITE; PS0235; UCH_2_3; 1.
CC Ub1 conjugation pathway; Hydrolase, Thiol protease; Multigene family.
FT ACT SITE 24 24 BY SIMILARITY.
FT ACT SITE 297 297 BY SIMILARITY.
FT ACT SITE 305 305 BY SIMILARITY.
SQ SEQUENCE 353 AA; 40591 MW; 4PFB39A225FE8F11 CRC64;

Query Match 9.4%; Score 383.5; DB 1; Length 353;
Best Local Similarity 30.3%; Pred. No. 5e-14;
Matches 104; Conservative 55; Mismatches 127; Indels 57; Gaps 12;

QY 111 AGLNLGNTCFANAALQCLTYTPPLANYMLSH-----EHSKTHA---EGFNMCTMQA 161
Db 15 AGLNLGNTCFNNSILOCLSNTRFLDYCLQRLYMRDLGHTSSANTALMEEF-----A 67
QY 162 HITQAL--SNPGDVTKMPFVINEMRIRAHFRNQEDAHFLOVTVDMQACGLNGSNK 219
Db 68 KLIQTIWTSPPNDVVSPEFTQIQRYAPRFGYNNQQAQEFRLFLDLGLH----NEVNR 123
QY 220 LDRHTQAT-----TLVCQIFGGLSRVVKLCKNGKGVSDTF 254
Db 124 VAARPKASPETLDLPDEBKQRMKRYLEREDSRIGDLFVGQLKSLTCTDCGYGVSTVF 183
QY 255 DFLDITLLEIK----AAQSVNKALEQFVKPEQUDGENSYKSKCKKMWVASKRFTIHRSS 310

```

```

Db 184 DPFWDLSLPIAKRGYPVTLMDCMRLFTKEDILDGDEKPTCCCRARAKCIKKFSVQRF 243
QY 311 NVLTSLSKRF--ANFTGGKIAKDVYP-EYLDIRPYMQNGEPIVYVLYAVLVHTGFNC 367
Db 244 KILVLHLKRFSEARIRASKLTTFNFPFLKDLDRFASQ-SCNHAAYNLVAVSNHSG-TT 301
QY 368 HAGHYFCYIKAS-NGLWYQMNDSIVSTDIRSVLSQAAVLFY 409
Db 302 MGGHYTAYCRSPVTGEWHTFNDSSVTPMSSSQVRTSDAYLLFY 344

RESULT 11
UBP8_SCHPO
ID UB8_SCHPO STANDARD; PRT; 449 AA.
AC Q09738;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Probable ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15)
DE (Ubiquitin thioesterase 8) (Ubiquitin-specific processing protease 8)
DE (Deubiquitinating enzyme 8).
GN UB8 OR SPAC13A11.04C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgroves J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Money P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fricz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Back A., Leinach H., Wambutt R., Purnelle B.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Lelaure V., Mottier S.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Hurst S.M.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Thode G.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Benito J.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RA "The genome sequence of Schizosaccharomyces pombe.";
RT Nature 415; 871-880(2002).
CC -!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -!- SIMILARITY: Belongs to peptidase family C19.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z54096; CAA90805.1; --
CC FIR; T37611; T37611.
CC MEROPS; C19.UPW; --

```

DR		GeneDB_Spombe; SPAC13A11.04c; -.
DR	InterPro;	IPIR001394; Peptidase_C19.
DR	InterPro;	IPRO01607; znf_UBP.
DR	Pfam;	PF00443; UCH; 1.
DR	Pfam;	PF02148; zf-UBP; 1.
DR	SMART;	SMO0290; ZNF_UBP; 1..
DR	PROSITE;	PS00972; UCH 2_1; 1.
DR	PROSITE;	PS00973; UCH 2_2; 1.
DR	PROSITE;	PS02353; UCH 2_3; 1.
KW	Ubl conjugation pathway; Hydrolase;	Thiol protease; Multigene family.
FT	ACT_SITE	154 154 BY SIMILARITY.
FT	ACT_SITE	387 387 BY SIMILARITY.
FT	ACT_SITE	395 395 BY SIMILARITY.
SQ	SEQUENCE	449 AA; 51720 MW; 281D26SD064CB3FC CRC64;
	Query Match	8.8%; Score 358.5; DB 1; Length 449;
	Best Local Similarity	29.6%; Pred. No. 1.5e-12;
	Matches 106; Conservative	62; Mismatches 143; Indels 47; Gaps 16;
QY	89	PQKVLFPSEIKLK-WOOTHRVG-----GLQLNGTCFANAALQCLTYTPEPLANYMLSH 142
DB	120	PEKY---NQMVLEAYRKPPVPCATAGLRGLGIONLGATCFMSVILOSILHNPVLNLFPSG 176
QY	143	EHSKTCHAEGFCMMCTWOAHITQALS--NPGDVTKPMFINEMRIARHLRFNGQEDAHE 200
DB	177	FHTSTDCRRPTCMTCADDMFSSIYNKNKSIFYGPTAVLNLMWKLSKSLCGYSQQDGHE 236
QY	201	FLOQTVDMOKACLNGSNKLDRTQA---TTLVG----QIFGGYLRSRYVKCLNCKGVSD 252
DB	237	FFVYLLQM-----HTESGGTSMPCTCIHRIFSGLNNVTCDKKERY 283
QY	253	TTFPYLDITLEIKAOSVNKALEOFVKPEQLDGENSYKCSKCKMWPASKRFTTHRSSNV 312
DB	284	AVDFPLMDISLDIN-EPTLQGCLERFSVEKV---QVSCHSCGSK-NAIQLYDFDKLPPT 337
QY	313	LTLSSLKEFA-NET-GGKIADVKYPDYLDIRPYMQSPNGEPIVVYLVVHTGFNCBA 369
DB	338	ICMQLKRFEQNFMAMSTKIDKQVSYPAFMRMYFNQ---DDVDYQLYSVVVCHKG-TLDT 393
QY	370	GHYFCYKIASGNLGYOWNDISIVSTSIRSLSQAAVFLFYIRSHDVXKNGGELTHPHS 427
DB	394	GHIYATIYYQN-QWFILDTTIVEVKESEVIN SQALLFY-HERQILIYSEMTEVKTEN 449
RESULT 12		
UB21_MOUSE		
ID	UB21_MOUSE	STANDARD; PRT; 566 AA.
AC	Q9ZLF6; Q9DOR1;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	Ubiquitin carboxyl-terminal hydrolase 21 (EC 3.1.2.15) (Ubiquitin thiolesterase 21) (Ubiquitin-specific processing protease 21) (deubiquitinating enzyme 21).	
GN	USP21 OR USP23.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutelestomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
RN	[1] TaxID=10090;	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RP	MEDLINE=20246307; PubMed=10786635;	
RA	Smith T.S., Southan C.;	
RT	"Sequencing, tissue distribution and chromosomal assignment of a novel ubiquitin-specific protease USP23";	
RRL	Biochim. Biophys. Acta 1490:184-188(2000).	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=C57BL/6J; TISSUE=Embryo;	
RC	MEDLINE=21085660; PubMed=11217851;	
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,	
RA	Arakawa T., Hara A., Fukunishi Y., Komoto H., Adachi J., Fukuda S.,	
RA	Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yananaka I.,	

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	257	YLDITILEIK-----AAQSVNKALEQFVKPEQLDGENSKCKKQWVPAKRFTHRS	310
CC	398	FCDLSIPFKGAGKVSRLDCFSLFTKEELESNPVCDRCRQKTRSTKLTQVQFP	457
CC	311	NVLTLKRPANFTGG--KIADVKYP-EYLDIRPMSQNGEPIVYVLYAVLVHVGNC	367
CC	458	RIIVLHLNRFSTRGSGIKSSVGVDFPQLRSLSGDFASDKAGSP-VYQLYALCNHSG-SV	515
CC	368	HAGHYFCYIKASNLWYQNDSTSVTSDIRSVLSQQAYVLFY	409
CC	516	HYGHYALCRQCTG-WHYNDSRVSPSENQVASSGCVLFY	556
CC	DB	UB46 HUMAN STANDARD; PRT; 366 AA.	
CC	AC	Q80V95; Q9H7U4; Q9H9T8;	
CC	DT	10-OCT-2003 (Rel. 42, Created)	
CC	DT	10-OCT-2003 (Rel. 42, Last sequence update)	
CC	DT	10-OCT-2003 (Rel. 42, Last annotation update)	
CC	DE	Ubiquitin carboxyl-terminal hydrolase 46 (EC 3.1.2.15) (Ubiquitin	
CC	DE	thiolesterase 46) (Ubiquitin-specific processing protease 46)	
CC	DE	(ubiquitin-activating enzyme 46).	
CC	GN	USP46.	
CC	OS	Homo sapiens (Human), and	
CC	OS	Mus musculus (Mouse).	
CC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
CC	OX	NCBI_TaxID=9606, 10090;	
CC	OX	[1]	
CC	RN	SEQUENCE FROM N.A.	
CC	RP	SPCIES=Human; TISSUE=Placenta, and Teratocarcinoma;	
CC	RC	Isegai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	
CC	RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,	
CC	RA	Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,	
CC	RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,	
CC	RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,	
CC	RA	Yanamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,	
CC	RA	Ninomiya K., Iwayanagi T.;	
CC	FT	"NEBO human cDNA sequencing project";	
CC	RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.	
CC	RL	[2]	
CC	RN	SEQUENCE FROM N.A.	
CC	RP	SPCIES=Human, and Mouse;	
CC	RC	STRAIN=FVB/N; TISSUE=Breast tumor, and Uterus;	
CC	EX	MEDLINE=22388257; PubMed=12477932;	
CC	RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,	
CC	RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,	
CC	RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,	
CC	RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,	
CC	RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,	
CC	RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,	
CC	RA	Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,	
CC	RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,	
CC	RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,	
CC	RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,	
CC	RA	Villalon D.K., Muzny N.C., Sodergren E.J., Lu X., Gibbs R.A.,	
CC	RA	Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,	
CC	RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,	
CC	RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,	
CC	RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,	
CC	RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,	
CC	RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;	
CC	RT	"Generation and initial analysis of more than 15,000 full-length	
CC	RT	human and mouse cDNA sequences.";	
CC	RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).	
CC	CC	-!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =	
CC	CC	ubiquitin + a thiol.	
CC	CC	-!- SIMILARITY: Belongs to peptidase family C19.	
CC	CC	-----	
CC	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	CC	the European Bioinformatics Institute. There are no restrictions on its	

CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
CC	EMBL; AK022614; BAB14133.1; -	
CC	EMBL; AK024318; BAB14881.1; -	
CC	EMBL; BC037574; AAH37574.1; -	
CC	EMBL; BC039916; AAH39916.1; -	
CC	MEROPS; C19.052; -	
CC	Genew; HGNC:20075; USP46.	
CC	InterPro; IPR001394; Peptidase_C19.	
CC	Pfam; PF00443; UCH_1.	
CC	PROSITE; PS00972; UCH_2_1; 1.	
CC	PROSITE; PS00973; UCH_2_2; 1.	
CC	PROSITE; PS00235; UCH_2_3; 1.	
CC	KW ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.	
CC	FT ACT_SITE 44 44	BY SIMILARITY.
CC	FT ACT_SITE 304 304	BY SIMILARITY.
CC	FT ACT_SITE 313 313	BY SIMILARITY.
CC	FT CONFLICT 250 250	I -> V (IN REF. 1; BAB14133).
CC	FT CONFLICT 265 265	H -> R (IN REF. 1; BAB14133).
CC	FT CONFLICT 265 265	H -> R (IN REF. 1; BAB14133).
CC	SQ SEQUENCE 366 AA; 42442 MW; 67BB113FC4081C46 CRC64;	
CC	Query Match 8.7%; Score 352.5; DB 1; Length 366;	
CC	Best Local Similarity 29.7%; Pred. No. 2.5e-12;	
CC	Matches 102; Conservative 55; Mismatches 125; Indels 61; Gaps 13;	
CC	112 GLQNLGNTCPANALQCLTTPPLANYLHSH-----EHSKTCHAGFCMCMCTMOAHIT 164	
CC	36 GLVNFNGNTCYCNSVLQALYFCRPFRENVLAYKAAQKKKKNLTLCLADLHFSIATQKKV- 94	
CC	165 QALSNPGDVTKPMFVINEMRIRARLFGN--QDAHEFLQYTVDM-----Q 210	
CC	95 -----GVIPPKKFI SRLK--ENDLFDNYMQDAHEFLNYLNTIADILQEEKQKQ 145	
CC	211 KACLNGSN--KLDHRTQATTLVCOIFGGYLRSVKCLNGKGVSDTFDPYLDITLAIKAA 267	
CC	146 NGKLKNGMNEPAEINKPELTWVHEIFQGLTTLNCTVSSKDEDFDLSDVDVEQN 205	
CC	268 QSVNKALEQFVKPEQLDGENSKCKKQWVPAKRFTHRSNNVLTSLAKFA----- 321	
CC	206 TSITHCLRDFSNTELTCLSEQKYCYETCCSKQEAQKRMVKKLPWLLALHLRFRKMYEQLH 265	
CC	322 NFGGKIADVKYPRYDIDIRPMSQNGEPI--VYVLYAVLVHVGNCAGHYFCYIKAS 379	
CC	266 RYT--KLRYRVFP--LELRLENTSSDAVNLDRMYDLVAVVHCGSGPNRGHYITIVK-S 320	
CC	380 NGLWYQNDSTSVS-----TSDIRSVLSQQAYVLFY 409	
CC	321 HGFLLPDDIVEKIDAKAIEEFYGLTSDI-SKNSSEGYILFY 362	
CC	-----	
CC	RESULT 14	
CC	UB21 HUMAN STANDARD; PRT; 565 AA.	
CC	AC Q9UK80; Q9FTV1; Q9NY4;	
CC	DT 16-OCT-2001 (Rel. 40, Created)	
CC	DT 16-OCT-2001 (Rel. 40, Last sequence update)	
CC	DT 10-OCT-2003 (Rel. 42, Last annotation update)	
CC	DE Ubiquitin carboxyl-terminal hydrolase 21 (EC 3.1.2.15) (Ubiquitin	
CC	thiolesterase 21) (Ubiquitin-specific processing protease 21)	
CC	DE (Deubiquitinating enzyme 21) (NEDD8-specific protease).	
CC	GN USP21 OR USP23	
CC	OS Homo sapiens (Human).	
CC	OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
CC	OX NCBI_TaxID=9606;	
CC	OX [1]	
CC	RP SEQUENCE FROM N.A.	
CC	RX MEDLINE=20246307; PubMed=10786635;	
CC	RA Smith T.S., Southan C.;	
CC	RT "Sequencing, tissue distribution and chromosomal assignment of a novel	

QY	23	SEAVSPGDM-----DAGSASWGAVSSINDVSNHTLSLGPVPGAVVSSSSVDPKSK	73
Db	111	SKSVSSGDLRPMGIALGHRGTGELGAALS-----RLALRPPTTIRSTSL-----	157
QY	74	PSPOKOALGDGIAPP-----OKVLFPFSKICLKWOQTHRVG	110
Db	158	-----RRLGGFCPPPTLFSIRTEPPASHGSFHMISARSSEPFYSDDKMA---HHTLLLG	208
QY	111	A----GLONLGNCFANAALQCLTYTPPLANYMLISHEHSKTCCHAEGFCMMCTMQ-AHITQA	166
Db	209	SGHVGLNLNGTCTFLNAVLQCLSTRPLRDFCLRRDPQREVPGGRAQELTEAFADVIGA	268
QY	167	LSNPG-DVIKPMFVINEMERIAHLRFFGNOEBAHEFLQTVIAM-----	209
Db	269	LWHPDSCSEAVNPTFRFAVFOKYVPFSGYSQDAQEFLKLMLERLHLEINRRGRAPPIL	328
QY	210	-----QKACLNGSNKLDRHQTQATTLVCOIFGGYILRSRVKCL	245
Db	329	ANGPVPSPPRRGALLPEELSDDDRANLMMWKYLERE---DSKIVDLFVGQLKSKLKCQ	385
QY	246	NKGVSDDTPPYLDITLEIK-----AAOSVNKALEQFVKPEQLDGENSKYKCKCKMWP	299
Db	386	ACGYRSTTTEVFCDLSLPKPKGFAGGKVSRLRDFCNLFKKEELESENAFPVCDRCQKTR	445
QY	300	ASKRFTIHRSSNVLTLSLKEFANFTGG---KIAKDVKYP-EYLDIRPMSQNPGEPIVYL	356
Db	446	STKKLIVQRPRLVLVHLNRFASRSIRKSSVGVDFPLQRLSLGDFASDKAGSP-VYQL	504
QY	357	YAVLVHTGFNGHAGHYFCYIKASGLWYQNDISVTSDIRSVLSQOAVVLFY	409
Db	505	YALCNEHG-SVHYGHYALCRQGTG-WHYVNDRSVPVSENQVASSGEGVLFY	555
RESULT 15			
ID	UBPA	YEAST	STANDARD; PRT; 792 AA.
AC	P53874;		
DT	01-OCT-1996	(Rel. 34, Created)	
DT	01-OCT-1996	(Rel. 34, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Ubiquitin carboxyl-terminal hydrolase 10 (EC 3.1.2.15) (Ubiquitin		
DE	thioesterase 10) (Ubiquitin-specific processing protease 10)		
DE	(Deubiquitinating enzyme 10) (Disrupter of telomere silencing protein		
DE	4).		
GN	UBP10 OR DOT4 OR YNL186W OR N1619.		
OS	Saccharomyces cerevisiae (Baker's yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.		
NCBI	TaxID=4932;		
RN	[1]_FUNCTION:		
RN	SEQUENCE FROM N.A.		
RP	Obermair B., Piravandi E., Rinke M., Domdey H.;		
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.		
RL	[2]		
RN	FUNCTION:		
RX	MEDLINE=98429517; PubMed=9755194;		
RA	Singer M.S., Kahana A., Wolf A.J., Meisinger L.L., Peterson S.E.,		
RA	Goggin C., Mahowald M., Gottschling D.E.;		
RT	"Identification of high-copy disruptors of telomeric silencing in		
RT	Saccharomyces cerevisiae.";		
RL	Genetics 150:613-632(1998).		
CC	!- FUNCTION: Has a role in telomere and HM loci silencing, which is		
CC	the repression of chromatin structure which leads to a stop in the		
CC	transcription of nearby genes.		
CC	!- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =		
CC	ubiquitin + a thiol.		
CC	!- SIMILARITY: Belongs to peptidase family C19.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		

CC diseases such as systemic lupus erythematosus, diseases of the skin such
CC as malignant melanoma, the bone marrow e.g. B- and T-lymphoid leukaemias,
CC the heart including myocardial infarction, red cells e.g. anaemia, the
CC thymus e.g. Hodgkin disease, B-cells e.g. peripheral B-cell neoplasms,
CC the kidney e.g. polycystic kidney disease, the breast including
CC periductal mastitis, the testis and epididymis e.g. syphilis, the
CC prostate e.g. nodular hyperplasia, the thyroid, e.g. hyperthyroidism, the
CC skeletal muscle e.g. rhabdomyosarcoma, the pancreas e.g. ectopic
CC pancreas, reduced platelet number e.g. HIV (human immunodeficiency virus)
CC -associated thrombocytopaenia and disorders involving precursor T-cell
CC neoplasms including precursor T lymphoblastic leukaemia/lymphoma
XX

XX SQ Sequence 762 AA;

Query Match 85.3%; Score 661; DB 4; Length 762;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPAYONQPGSEAVSPGDMAGSAGVSSLDVSNHTLSLGPVPGA 60

Db 1 MTIVDKASSSDPAYONQPGSEAVSPGDMAGSAGVSSLDVSNHTLSLGPVPGA 60

QY 61 VVYSSSVDPKSPKQDQALGDGIAPPKQVLPFSEKICLKWQOQTHRVGAGLQNLGNTC 120

Db 61 VVYSSSVDPKSPKQDQALGDGIAPPKQVLPFSEKICLKWQOQTHRVGAGLQNLGNTC 120

QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFCMCTQOAHITQALSNPDGVDIKMFVI 180

Db 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFCMCTQOAHITQALSNPDGVDIKMFVI 180

QY 181 NEMRRTARHLRFQNGDEAHEFLQYTDAMQACINGSNKLDRTQATTLVCOIFGGVYLS 240

Db 181 NEMRRTARHLRFQNGDEAHEFLQYTDAMQACINGSNKLDRTQATTLVCOIFGGVYLS 240

QY 241 RVKCLCKGVSDFDPVLDITLIEKAAQSNKALEQFVKPEQLDGENSYKSCCKKWVPA 300

Db 241 RVKCLCKGVSDFDPVLDITLIEKAAQSNKALEQFVKPEQLDGENSYKSCCKKWVPA 300

QY 301 SKRFTIHRSSNVLTSLKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPFIVVLYAVL 360

Db 301 SKRFTIHRSSNVLTSLKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPFIVVLYAVL 360

QY 361 VHTGFNCHAGHYCYIKASNGVLQYQNDMSIVSTDSVLSQQAAYLVFIYRSHDVKNNGE 420

Db 361 VHTGFNCHAGHYCYIKASNGVLQYQNDMSIVSTDSVLSQQAAYLVFIYRSHDVKNNGE 420

QY 421 LTHPTSPGQSSPRPVTSQRVVTNKQAAPGFIGPQLPSHMIKNPPHNLGTGPKLDTFSS 480

Db 421 LTHPTSPGQSSPRPVTSQRVVTNKQAAPGFIGPQLPSHMIKNPPHNLGTGPKLDTFSS 480

QY 481 MSSPNGNSVNRASPVNASQVQNSVNRSSVIPHEPKQKITITISHNKLPVRCQSQPN 540

Db 481 MSSPNGNSVNRASPVNASQVQNSVNRSSVIPHEPKQKITITISHNKLPVRCQSQPN 540

QY 541 LHSNLENPKVPSSITINSVOSTSNASTMSVSSVTKPIPRSECSQVPMNGSKLN 600

Db 541 LHSNLENPKVPSSITINSVOSTSNASTMSVSSVTKPIPRSECSQVPMNGSKLN 600

QY 601 SSVLPVYGASSEDSDSESKGLKXGNGITTVSGHSPGQADEATPHEIQEPMTLNGA 660

Db 601 SSVLPVYGASSEDSDSESKGLKXGNGITTVSGHSPGQADEATPHEIQEPMTLNGA 660

QY 661 NSADSDDPKENGLAPDGASQCGQFALHSENPFAKNGLPKMLPAPLISPEDKILETF 720

Db 661 NSADSDDPKENGLAPDGASQCGQFALHSENPFAKNGLPKMLPAPLISPEDKILETF 720

QY 721 RLSNKLKGTDMSPAGAEPPEDRDAEPQGPSAAESLEE 762

Db 721 RLSNKLKGTDMSPAGAEPPEDRDAEPQGPSAAESLEE 762

ID	ABU08951 standard; prote.n; 762 AA.
XX	
AC	ABU08951;
XX	
DT	10-JUN-2003 (first entry)
XX	
DE	Human ubiquitin protease.
XX	
KW	Human; enzyme; ubiquitin protease; spleen disorder; Hodgkin's disease;
KW	lung disorder; adult respiratory distress syndrome; colon disorder;
KW	inflammatory bowel disease; liver disorder; jaundice; uterine disorder;
KW	endometriosis; brain disorder; Alzheimer's disease; T-cell disorder;
KW	acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;
KW	heart disorder; ischaemic heart disease; blood vessel disorder;
KW	atherosclerosis; red blood cell disorder; anaemia; thymus disorder;
KW	DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder;
KW	polycystic kidney disease; glomerulonephritis; breast disorder; mastitis;
KW	testicular disorder; sexually transmitted disease; thyroid disorder;
KW	hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;
XX	whipple disease; tumour; cancer.
OS	Homo sapiens.
XX	
PN	US2003037350-A1.
XX	
PD	20-FEB-2003.
XX	
PF	05-JUN-2002; 2002US-00163547.
XX	
PR	29-SEP-1999; 99US-00407356.
PR	05-NOV-1999; 99US-00435311.
PR	01-FEB-2000; 2000US-00496005.
PR	11-FEB-2000; 2000US-0182009P.
PR	14-FEB-2000; 2000US-0182408P.
PR	28-FEB-2000; 2000US-0185503P.
PR	02-NOV-2000; 2000US-00704918.
PR	12-FEB-2001; 2001US-00781598.
PR	14-FEB-2001; 2001US-00782952.
PR	28-FEB-2001; 2001US-00796100.
XX	(MILL-) MILLENNIUM PHARM INC.
XX	
PI	Glucksmann MA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;
XX	
DR	WPI: 2003-342047/32.
XX	N-PSDB; ABX33880.
PT	
XX	New nucleic acid or polypeptide, useful for preparing a composition for
XX	diagnosing or treating diseases e.g., tumor.
PS	Claim 9; Fig 1; 281pp; English.
XX	
CC	The invention relates to a new isolated nucleic acid molecule encoding
CC	one of 6 polypeptides (ubiquitin protease, lipase, Dynamain, short chain
CC	dehydrogenase, ADAM-17 (a disintegrin and metalloprotease domain protein
CC	with threonase, ADAM-17 (TS) domains) and gamma butyrobetaine-hydroxylase
CC	(gamma BH). Also included are the polypeptide, host cells containing the
CC	nucleic acids, an antibody that selectively binds to the polypeptide, a
CC	method for producing the polypeptide, a method for detecting the presence
CC	of the polypeptide or the nucleic acid in a sample, a method for
CC	identifying a compound that binds to the polypeptide, a method for
CC	modulating the activity of the polypeptide and a method for identifying a
CC	compound that modulates the activity of the polypeptide. The nucleic acid
CC	or polypeptide is useful for preparing a composition for diagnosing or
CC	treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's
CC	disease), lung disorders (e.g. adult respiratory distress syndrome,
CC	pulmonary oedema, chronic bronchitis and emphysema), colon disorders
CC	(e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),
CC	liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver
CC	disease), uterine and endometrial disorders (e.g. endometriosis and
CC	menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's
CC	disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell
CC	disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders

CC (e.g. urticaria, dermatitis and lupus erythematosus), heart disorders
CC (e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),
CC blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and
CC Raynaud disease), red blood cell disorders (e.g. anaemia), thymus
CC disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),
CC kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),
CC breast disorders (e.g. mastitis), testicular disorders (e.g. sexually
CC transmitted diseases and cryptorchidism), thyroid disorders (e.g.
CC hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal
CC disorders (e.g. whipple disease), as well as tumours and cancers of the
CC above listed organs/cells. Many more diseases and disorders are listed in
CC the specification. The present sequence represents human ubiquitin
XX protease
SQ Sequence 762 AA;

Query Match		85.3%;	Score 661;	DB 6;	Length 762;
Best Local Similarity		99.9%;	Pred. No. 0;		
Matches 761;		Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
QY	1	MTIVDKASESSDPSAYONQPGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA	60		
DB	1	MTIVDKASESSDPSAYONQPGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA	60		
QY	61	VYSSSSVDPKSPQKQALGDGIAPPOKVLFPSEKICLKQOQTHRVGAGLQNLGNTC	120		
DB	61	VYSSSSVDPKSPQKQALGDGIAPPOKVLFPSEKICLKQOQTHRVGAGLQNLGNTC	120		
QY	121	FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTQOAHITQALSNGPDVVKPMFVI	180		
DB	121	FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTQOAHITQALSNGPDVVKPMFVI	180		
QY	181	NEMRIARHLRFGNORDAHEFLQYTVDMOKACLNGSNKLDRTQATTLVCOIFGGYLS	240		
DB	181	NEMRIARHLRFGNORDAHEFLQYTVDMOKACLNGSNKLDRTQATTLVCOIFGGYLS	240		
QY	241	RVKCLNCKGVSTFDPYLDITLKAQSVNKALEQFVKPEQLDGENSYKSKCKKWPA	300		
DB	241	RVKCLNCKGVSTFDPYLDITLKAQSVNKALEQFVKPEQLDGENSYKSKCKKWPA	300		
QY	301	SKRFTIHRSSNVLTLSLKRFANFTGKTAKDYKPYEYLDIRPYMGPNQGEPIVYVLYAVL	360		
DB	301	SKRFTIHRSSNVLTLSLKRFANFTGKTAKDYKPYEYLDIRPYMGPNQGEPIVYVLYAVL	360		
QY	361	VHTGNCCHAGHYFCYIKASNGLWYQWNSIVSTDIRSVLSQAVYLFIRSHDVKNGE	420		
DB	361	VHTGNCCHAGHYFCYIKASNGLWYQWNSIVSTDIRSVLSQAVYLFIRSHDVKNGE	420		
QY	421	LTHPTSPGQSPRPVTSQVVVTKQAAPGFTGPOLPSHMIKNPPLHNGTGPKDTPSSS	480		
DB	421	LTHPTSPGQSPRPVTSQVVVTKQAAPGFTGPOLPSHMIKNPPLHNGTGPKDTPSSS	480		
QY	481	MSSPNGNSVNRASPVNASVQWNSVNRSSVIEHPKPKQKITTISHNKLPVRCQSQPN	540		
DB	481	MSSPNGNSVNRASPVNASVQWNSVNRSSVIEHPKPKQKITTISHNKLPVRCQSQPN	540		
QY	541	LHNSLENPTKVPSSITTSNAVOSTSNASTWSVSKVTKEIPRSESCSQPVWNGSKLN	600		
DB	541	LHNSLENPTKVPSSITTSNAVOSTSNASTWSVSKVTKEIPRSESCSQPVWNGSKLN	600		
QY	601	SSVLVPYGAESSESDSESKGLGKENGIGITTVSSHSPQDAEDEATPHELQEPWTLNGA	660		
DB	601	SSVLVPYGAESSESDSESKGLGKENGIGITTVSSHSPQDAEDEATPHELQEPWTLNGA	660		
QY	661	NSADSDSDPKENGLAPDGCACQOPALHSENPPAKANGLPKMLPALLSPEDKILETF	720		
DB	661	NSADSDSDPKENGLAPDGCACQOPALHSENPPAKANGLPKMLPALLSPEDKILETF	720		
QY	721	RLSNKLGSTDEMSAPGAERFPEDRDAEPQPGSPAESLEE	762		
DB	721	RLSNKLGSTDEMSAPGAERFPEDRDAEPQPGSPAESLEE	762		

RESULT 5
AAB95146
ID AAB95146 standard; protein; 1197 AA.

XX AAB95146;
AC
DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:17169.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX Homo sapiens.

XX EP1074617-A2.
XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.
XX 27-AUG-1999; 99JP-00300253.
XX 11-JAN-2000; 2000JP-00118776.
XX 02-MAY-2000; 2000JP-00183767.
XX 03-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 17169; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 1197 AA;

Query Match 72.9%; Score 565; DB 4; Length 1197;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 765; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASESSDPSAYONQPGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASESSDPSAYONQPGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60

QY 61 VYSSSSVDPKSKPSPQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VYSSSSVDPKSKPSPQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTMQAHTOALSNGPGVVKPMFVI 180
DB 121 FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTMQAHTOALSNGPGVVKPMFVI 180
QY 181 NEMRIARHLRFGNQEADAEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
DB 181 NEMRIARHLRFGNQEADAEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
QY 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGNSYKSCCKKWPA 300
DB 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGNSYKSCCKKWPA 300
QY 301 SKRFTTHRSSNVLTLSLRFANFTGKIAKOVKPYEYLDIRPYMSQNGEPIVYVLYAVL 360
DB 301 SKRFTTHRSSNVLTLSLRFANFTGKIAKOVKPYEYLDIRPYMSQNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQOAYVLYFIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQOAYVLYFIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQIRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480
DB 421 LTHPTSPGQSSPRPVISQIRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480
QY 481 MSSPFGNSSVNRASPVNASQVNSVSVIPEHPKOKITISHNKLPVRCQSQPN 540
DB 481 MSSPFGNSSVNRASPVNASQVNSVSVIPEHPKOKITISHNKLPVRCQSQPN 540
QY 541 LHSNLSNPKVPSSITNSAVOSTSNASTMSVSSKVTPIPRSESCSPVWNGSKLIN 600
DB 541 LHSNLSNPKVPSSITNSAVOSTSNASTMSVSSKVTPIPRSESCSPVWNGSKLIN 600
QY 601 SSVLPVYGAESSEDSDESKGLKENGIGITVSSHSPQDAEDAEAPHLOEPWTLNGA 660
DB 601 SSVLPVYGAESSEDSDESKGLKENGIGITVSSHSPQDAEDAEAPHLOEPWTLNGA 660
QY 661 NSADSDPKENGLAPDGASCOGAPALHSENPFKAKNGLPKLMPAPLLSLPDKILETF 720
DB 661 NSADSDPKENGLAPDGASCOGAPALHSENPFKAKNGLPKLMPAPLLSLPDKILETF 720
QY 721 RLSNKLKGSTDEMSAPGAERPPEDRAEPQPGSPAESLEPDAEA 767
DB 721 RLSNKLKGSTDEMSAPGAERPPEDRAEPQPGSPAESLEPDAEA 767
RESULT 6
ID ABP69291 standard; protein; 1270 AA.
XX AC ABP69291;
XX DE Human polypeptide SEQ ID NO 1338.
XX 20-JAN-2003 (first entry)
XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
XX cell-proliferative disorder; neurodegenerative disease; bacterial;
XX Parkinson's disease; Alzheimer's disease; autoimmune disease;
XX multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
XX arthritis; cytostatic; immunomodulator; nontropic; neuroprotective;
XX antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
XX haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
XX antiarthritic.
XX Homo sapiens.
XX WO2002/0539-A2.
XX 12-SEP-2002.
PD

XX 05-MAR-2002; 2002WO-US0005095.
XX 05-MAR-2001; 2001US-00799451.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
XX Wehrman T, Wang J, Wang D, Drmanac RT;
XX WPI; 2002-759812/82.
XX N-PSDB; AB211508.
XX New polynucleotides comprising sequences assembled from expressed
XX sequence tags (ESTs), useful for treating cell-proliferative,
XX neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
XX or coagulation disorders.
XX Claim 9; SEQ ID NO 1338; 1012pp + Sequence Listing; English.
XX The invention relates to an isolated polynucleotide (I) comprising a
XX nucleotide sequence selected from any of 948 sequences (AB211119-
XX AB212066) or their mature protein coding portion, active domain coding
XX protein or complementary sequences. The polynucleotides are useful for
XX identifying expressed genes or for physical mapping of human genome. The
XX encoded polypeptides (ABP6892-ABP6984) are useful as molecular weight
XX markers, as a food supplement, for generating antibodies, in medical
XX imaging, screening and diagnostic assays and for treating cell-
XX proliferative disorders (cancer), autoimmune diseases (multiple sclerosis,
XX or Alzheimer's disease), genetic disorders, myeloid or lymphoid disorders,
XX diabetes, lupus) genetic disorders, wound, burns, incision, ulcers, liver
XX platelet or coagulation disorders (bacterial, viral, fungal, parasitic),
XX or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
XX arthritis, etc. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1270 AA;
QY Query Match 72 9%; Score 565; DB 5; Length 1270;
DB Best Local Similarity 99.7%; Pred. No. 0;
Matches 765; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MTIVDKASESDPSAYQNPQGSSEAVSPGMDAGSAGVSSLVNDVSNHTLSLGPVGA 60
DB 1 MTIVDKASESDPSAYQNPQGSSEAVSPGMDAGSAGVSSLVNDVSNHTLSLGPVGA 60
QY 61 VYSSSSVDPKSKPSPQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
DB 61 VYSSSSVDPKSKPSPQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
QY 121 FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTMQAHTOALSNGPGVVKPMFVI 180
DB 121 FANAALQCLTYTPPLANYMLSHHSKTCCHAEFCMCTMQAHTOALSNGPGVVKPMFVI 180
QY 181 NEMRIARHLRFGNQEADAEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
DB 181 NEMRIARHLRFGNQEADAEFLQYTVDMQKACLSNGSKLDRHTQATTLVCOIFGGLYLS 240
QY 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGNSYKSCCKKWPA 300
DB 241 RVKCLNCKGVSTDFPYLDITLLEIKAAQSVNKALEQFVKPEQLDGNSYKSCCKKWPA 300
QY 301 SKRFTTHRSSNVLTLSLRFANFTGKIAKOVKPYEYLDIRPYMSQNGEPIVYVLYAVL 360
DB 301 SKRFTTHRSSNVLTLSLRFANFTGKIAKOVKPYEYLDIRPYMSQNGEPIVYVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQOAYVLYFIRSHDVKNNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVLSQOAYVLYFIRSHDVKNNGE 420
QY 421 LTHPTSPGQSSPRPVISQIRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS 480

Db 421 LTHPTSPGSSPRPVISQRTVNTKQAPGFIGQLPSHMIKNPPLHNGTGELKDTSSS 480
 QY 481 MSSPNGSSVNRASPVNASVQVNSVNRSSVPEHPKQKQITISIHNLKLPVRCQSQPN 540
 Db 481 MSSPNGSSVNRASPVNASVQVNSVNRSSVPEHPKQKQITISIHNLKLPVRCQSQPN 540
 QY 541 LHSNLENPTKPVPSSTITNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLN 600
 Db 541 LHSNLENPTKPVPSSTITNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLN 600
 QY 601 SSVLVYVGAESSEDSDEESKGLKENGIGTIVSSHSPQDAEATPHELQEPMTLNGA 660
 Db 601 SSVLVYVGAESSEDSDEESKGLKENGIGTIVSSHSPQDAEATPHELQEPMTLNGA 660
 QY 661 NSADSDSDPKENGLAPDGAQSQQAPALHSENPEAKANGLPGKLPAPLLSLPDKILETF 720
 Db 661 NSADSDSDPKENGLAPDGAQSQQAPALHSENPEAKANGLPGKLPAPLLSLPDKILETF 720
 QY 721 RLNSKLKSTDEMSAPGAERGGPDRDAEPQGPSAAESLEEDAAA 767
 Db 721 RLNSKLKSTDEMSAPGAERGGPDRDAEPQGPSAAESLEEDAAA 767

RESULT 7

ABB06118
 ID ABB06118 standard; protein; 558 AA.

XX ABB06118;

DT 10-MAY-2002 (first entry)

XX Human NS protein sequence SEQ ID NO:210.

DE Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

XX Human; cytostatic; osteopathic; gynaecological; neuroprotective;

CC ABL39691 to ABL39818 represent novel human nucleic acid sequences
 CC encoding the proteins given in ABB06037 to ABB06164. The novel sequences
 CC (NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,
 CC antirheumatic, antiarthritic, antipsoriatic, ophthalmological, virucide,
 CC vasotropic, antiarteriosclerotic, antiinflammatory, dermatological,
 CC anorectic, muscular, anti-HIV, antinfertility, cardiovascular,
 CC anticoagulant, antifibrinolytic, hypotension, antiasthmatic, cardiant,
 CC immunomodulator, anticonvulsant, antidiabetic, tranquilliser, antiulcer,
 CC antidepressant, gastrointestinal, neuroleptic, cerebroprotective,
 CC neotropic and contraceptive activities. The NS can be used in vaccines,
 CC gene therapy and antitense therapy. Nucleic acids, expression vectors and
 CC antibodies from the present invention can be used for treating and
 CC diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative
 CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
 CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
 CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
 CC disease, coagulation disease, ischaemia, hypertension, asthma, immune
 CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
 CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
 CC Alzheimer's disease and as a contraceptive
 XX
 SQ Sequence 558 AA;

Query Match 57.8%; Score 448; DB 5; Length 558;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 448; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 319 RFANFTGGKIAKDVKYPEYLDIRPYMSQNGPPIVYLVAVLVTGFCNCHAGHYCYIKA 378

Db 111 RFANFTGGKIAKDVKYPEYLDIRPYMSQNGPPIVYLVAVLVTGFCNCHAGHYCYIKA 170

QY 379 SNGLMYQMNDSIVSTDIRSVLSQAYVLFYIRSHDVKNKGELTHTHSPGQSSRPVTS 438

Db 171 SNGLMYQMNDSIVSTDIRSVLSQAYVLFYIRSHDVKNKGELTHTHSPGQSSRPVTS 230

QY 439 QRVTNKAQAPGFIGQLPSHMIKNPPLHNGTGPLKDTSPSSMSSPNGSNVNRASPVNA 498

Db 231 QRVTNKAQAPGFIGQLPSHMIKNPPLHNGTGPLKDTSPSSMSSPNGSNVNRASPVNA 290

QY 499 SASVQVNSVNRSSVPEHPKQKITISIHNLKLPVRCQSQPVNMGSKLNSVLPVYGAESSEDSDE 558

Db 291 SASVQVNSVNRSSVPEHPKQKITISIHNLKLPVRCQSQPVNMGSKLNSVLPVYGAESSEDSDE 350

QY 559 TNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLNSVLPVYGAESSEDSDE 618

Db 351 TNSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLNSVLPVYGAESSEDSDE 410

QY 619 SKGLKENGIGTIVSSHSPQDAEATPHELQEPMTLNGANSADSDPKENGLAPDG 678

Db 411 SKGLKENGIGTIVSSHSPQDAEATPHELQEPMTLNGANSADSDPKENGLAPDG 470

QY 679 ASCQGPALHSENFPAKANGLPGLMPAPLLSLPDKILETFRLSNKLKSTDEMSAPGA 738

Db 471 ASCQGPALHSENFPAKANGLPGLMPAPLLSLPDKILETFRLSNKLKSTDEMSAPGA 530

QY 739 ERGPPEDRDAEPQGPSAAESLEEDAAA 766

Db 531 ERGPPEDRDAEPQGPSAAESLEEDAAA 558

RESULT 8

AAU23207

ID AAU23207 standard; protein; 156 AA.

XX AAU23207;

AC AAU23207;

XX 17-DEC-2001 (first entry)

DT 17-DEC-2001 (first entry)

XX Novel human enzyme polypeptide #293.

DE Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;

Wed Aug 18 13:53:03 2004

us-10-049-745-4.oli10.rag

KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.

OS Homo sapiens.

PN WO20015301-A2.

XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US001239.

XX 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216847P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.

PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236127P.
PR 29-SEP-2000; 2000US-0236167P.
PR 29-SEP-2000; 2000US-0236168P.
PR 29-SEP-2000; 2000US-0236169P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236302P.
PR 02-OCT-2000; 2000US-0237337P.
PR 02-OCT-2000; 2000US-0237338P.
PR 02-OCT-2000; 2000US-0237339P.
PR 02-OCT-2000; 2000US-0237340P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0246178P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250161P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.

PR	11-DEC-2000; 2000US-0254097P.
PR	05-JAN-2001; 2001US-0259678P.
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	Rosen CA, Barash SC, Ruben SM;
XX	WPI; 2001-465566/50.
DR	N-PSDB; AAS41077.
XX	
PT	Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT	treating neural, immune system, muscular, reproductive, pulmonary,
PT	cardiovascular, renal, proliferative disorders and cancerous diseases.
XX	
PS	Claim 11; SEQ ID NO 1203; 1180pp; English.
XX	
CC	The present invention relates to the isolation of novel human enzyme
CC	polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC	encoding them. The enzyme polypeptides of the invention may comprise the
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC	isomerases or ligases. The sequences of the invention are useful in the
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of
CC	disorders including hyperproliferative disorders (e.g. cancer),
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC	arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC	disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC	cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC	(e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC	infectious disorders (e.g. Influenza). The polynucleotides of the
CC	invention can also be used in gene therapy. AAU22915-AAU23814 represent
CC	the novel human enzyme polypeptides of the invention. Note: The sequence
CC	data for this patent did not form part of the printed specification, but
CC	was obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 156 AA;

Query Match 19.1%; Score 148; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 6.7e-141;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY	1	MTIIVDKASESDPAYQNQPSSSEAVSPGDM DAGSAGVSSLNDVSNHTLSLGPVPGA	60
Dd	2	MTIIVDKASESDPAYQNQPSSSEAVSPGDM DAGSAGVSSLNDVSNHTLSLGPVPGA	61
QY	61	VVYSSTSVPDKSKTSPQKDQALGDGIAPPKVLPESEKICLKWOOTHRVGAGLQNLNTC	120
Dd	62	VVYSSTSVPDKSKTSPQKDQALGDGIAPPKVLPESEKICLKWOOTHRVGAGLQNLNTC	121
QY	121	FANAALOCLTYTPPLANMLSHESKTC	148
Dd	122	FANAALOCLTYTPPLANMLSHESKTC	149

RESULT 9
AAU23743
ID AAU23743 standard; protein; 157 AA.
XX
AC AAU23743;
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #829.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
OS Homo sapiens.
XX

XX WPI: 2001-465566/50.
DR N-PSDB; AAS41613.
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
PT treating neural, immune system, muscular, reproductive, pulmonary,
PT cardiovascular, renal, proliferative disorders and cancerous diseases.
XX Claim 11; SEQ ID NO 1739, 1180pp; English.
XX The present invention relates to the isolation of novel human enzyme
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
CC encoding them. The enzyme polypeptides of the invention may comprise the
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
CC isomerases or ligases. The sequences of the invention are useful in the
CC diagnosis, treatment, prevention and/or prognosis of a wide range of
CC disorders including hyperproliferative disorders (e.g. cancer),
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
CC arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
CC disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
CC cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
CC (e.g. haemophilia), reproductive disorders (e.g. infertility) and
CC infectious disorders (e.g. influenza). The polynucleotides of the
CC invention can also be used in gene therapy. AAU2915-AAU23814 represent
CC the novel human enzyme polypeptides of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX Sequence 157 AA;
SQ
Query Match 16.5%; Score 128; DB 4; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.2e-120;
Matches 128; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MTIVDKASESSDPSAYQNPQCSSEAVSPGMDAGSASWGAIVSSINDVSNHTLSIGPVPGA 60
Db 2 MTIVDKASESSDPSAYQNPQCSSEAVSPGMDAGSASWGAIVSSINDVSNHTLSIGPVPGA 61
Qy 61 VVYSSSSVPDKSKPEPOKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
Db 62 VVYSSSSVPDKSKPEPOKQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 121
Qy 121 FANAALQC 128
Db 122 FANAALQC 129
RESULT 10
AAO12240
ID AAO12240 standard; protein; 125 AA.
XX AAO12240;
AC AAO12240;
XX 06-NOV-2001 (first entry)
XX Human polypeptide SEQ ID NO 26132.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX Homo sapiens.
OS WO200164835-A2.
XX 07-SEP-2001.
XX 26-FEB-2001; 2001WO-US004927.
XX 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.

PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239337P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PR Rosen CA, Barash SC, Ruben SM;
XX


```
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX N-PSDB; AAI92171.
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX and treating e.g. leukemia, inflammation and immune disorders.
XX Claim 20; SEQ ID NO 26132; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 125 AA;
XX
XX Query Match 16.0%; Score 124; DB 4; Length 125;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-116;
XX Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MTIVDKAESDPSAYQONQPGSSEAVSPGMDAGASWGAVALNDVSNHTLSLGPVGA 60
XX |
XX DB 2 MTIVDKAESDPSAYQONQPGSSEAVSPGMDAGASWGAVALNDVSNHTLSLGPVGA 61
XX |
XX QY 61 VYSSSSVPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQTHRVGAGLQNGTC 120
XX |
XX DB 62 VYSSSSVPDKSPQKQDQALGDGIAPPQKVLFPSEKICLKWQTHRVGAGLQNGTC 121
XX |
XX QY 121 FANA 124
XX |
XX DB 122 FANA 125
XX |
XX
XX RESULT 11
XX AAM16544
XX ID AAM16544 standard; protein; 112 AA.
XX AC AAM16544;
XX
XX DT 12-OCT-2001 (first entry)
XX
XX DE Peptide #2978 encoded by probe for measuring cervical gene expression.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer.
XX OS Homo sapiens.
XX
XX PN WO200157278-A2.
XX
XX Q9-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000670.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
```

```
PR 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 27; SEQ ID NO 21370; 487pp; English.
XX
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX can be used to produce a single exon microarray, which can be used for
XX measuring human gene expression in a sample derived from human cervical
XX epithelial cells. By measuring gene expression, the probes are therefore
XX useful in grading and/or staging of diseases of the cervix, notably
XX cervical cancer. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 112 AA;
XX
XX Query Match 14.5%; Score 112; DB 4; Length 112;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-104;
XX Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 591 PYMNGKSLNSVLVPYGAESSEDSDEESKGLKENGIGTIVSSHSPGQDADEEATPHE 650
XX |
XX DB 1 PYMNGKSLNSVLVPYGAESSEDSDEESKGLKENGIGTIVSSHSPGQDADEEATPHE 60
XX |
XX QY 651 LOEPMTLGANSADSDPKENGLAPDGCSCQGPALHSENPFKANGLPKG 702
XX |
XX DB 61 LOEPMTLGANSADSDPKENGLAPDGCSCQGPALHSENPFKANGLPKG 112
XX |
XX
XX RESULT 12
XX ABB35528
XX ID ABB35528 standard; peptide; 112 AA.
XX AC ABB35528;
XX
XX DT 04-FEB-2002 (first entry)
XX
XX DE Peptide #3034 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX
XX PN WO200157277-A2.
XX
XX PD 09-AUG-2001.
XX
XX PF 30-JAN-2001; 2001WO-US000669.
XX
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX
XX FA (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
```

PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human fetal liver.

XX Claim 27; SEQ ID NO 28163; 639pp + Sequence Listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 PVMNGSKLNSSVLPVYGAESEDSDEESKGLKENGIGTIVSSHSPGQDADEEATPHE 650
Db 1 PVMNGSKLNSSVLPVYGAESEDSDEESKGLKENGIGTIVSSHSPGQDADEEATPHE 60

QY 651 LQEPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFAKANGLPK 702

Db 61 LQEPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFAKANGLPK 112

RESULT 13

ABB30354
ID ABB30354 standard; peptide; 112 AA.

AC ABB30354;

XX 01-FEB-2002 (first entry)

XX Peptide #3005 encoded by breast cell single exon nucleic acid probe.

XX Human; microarray; single exon probe; gene expression; breast; disease;
XX cancer.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes, useful
XX for measuring gene expression in sample derived from human breast,
XX comprises number of single exon nucleic acid probes.

XX Claim 27; SEQ ID NO 13322; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human breast and BT 474 cells. The method involves contacting the

CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a peptide
CC encoded by a single exon nucleic acid probe of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 591 PVMNGSKLNSSVLPVYGAESEDSDEESKGLKENGIGTIVSSHSPGQDADEEATPHE 650
Db 1 PVMNGSKLNSSVLPVYGAESEDSDEESKGLKENGIGTIVSSHSPGQDADEEATPHE 60

QY 651 LQEPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFAKANGLPK 702

Db 61 LQEPMTLNGANSADSDPKENGLAPDGASCOGQOPALHSENPFAKANGLPK 112

RESULT 14

AAM68717
ID AAM68717 standard; protein; 112 AA.

AC AAM68717;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 29023.

XX Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.

XX Example 4; SEQ ID NO 29023; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human

CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention

XX SQ Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
 Best Local Similarity 100.0%; Pred. No. 1.4e-104;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 591 PVWNGSKLNSVLVPYGAESSESDSESKGLGKENGIGITIVSSHSPQDAEAEATPHE 650
 Db 1 PVWNGSKLNSVLVPYGAESSESDSESKGLGKENGIGITIVSSHSPQDAEAEATPHE 60
 QY 651 LQEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPPAKANGLFGK 702
 Db 61 LQEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPPAKANGLFGK 112

RESULT 15

AAM04259
 ID AAM04259 standard; protein; 112 AA.

XX AC AAM04259;

XX DT 09-OCT-2001 (first entry)

XX DE Peptide #2941 encoded by probe for measuring breast gene expression.

XX KW Probe; human; breast disease; breast cancer; development disorder;

XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.

XX OS Homo sapiens.

XX PN WO200157270-A2.

XX PD 09-AUG-2001.

XX PF 29-JAN-2001; 2001WO-US000661.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-476286/51.

XX PT Novel single exon nucleic acid probe used to measuring gene expression in
 XX a human breast.

XX PS Claim 27; SEQ ID NO 12999; 322pp; English.

CC The present invention relates to novel single exon nucleic acid probes
 CC (see AAI00010-AA110067). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for measuring human gene expression in
 CC a human breast sample, where the probe hybridises at high stringency to a
 CC nucleic acid expressed in the human breast. The probes are useful for
 CC predicting, diagnosing, grading, staging, monitoring and prognosing
 CC diseases of the human breast, particularly those diseases with polygenic
 CC aetiology. The diseases include: breast cancer, disorders of polycystic
 CC inflammatory diseases of the breast, fibrocystic changes, proliferative
 CC breast disease and non-carcinoma tumours. Note: the sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 112 AA;

Query Match 14.5%; Score 112; DB 4; Length 112;
 Best Local Similarity 100.0%; Pred. No. 1.4e-104;
 Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 591 PVWNGSKLNSVLVPYGAESSESDSESKGLGKENGIGITIVSSHSPQDAEAEATPHE 650
 Db 1 PVWNGSKLNSVLVPYGAESSESDSESKGLGKENGIGITIVSSHSPQDAEAEATPHE 60
 QY 651 LQEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPPAKANGLFGK 702
 Db 61 LQEPMTLNGANSADSDPKENGLAPDAGSCQGPALHSENPPAKANGLFGK 112

RESULT 16

ABG38302
 ID ABG38302 standard; peptide; 112 AA.

XX AC ABG38302;

XX DT 19-AUG-2002 (first entry)

XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 27967.

XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;

XX KW chronic obstructive pulmonary disease; interstitial lung disease;

XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;

XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;

XX KW primary ciliary dyskinesia; pulmonary hypertension;

XX KW hyaline membrane disease.

XX OS Homo sapiens.

XX PN WO200186003-A2.

XX PD 15-NOV-2001.

XX PF 30-JAN-2001; 2001WO-US000665.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX PT Spatially-addressable set of single exon nucleic acid probes, used to
 XX measure gene expression in human lung samples.

XX PS Claim 27; SEQ ID NO 27967; 634pp; English.

CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a sample
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung

mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, pulmonary haemorrhage, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 14.5%; Score 112; DB 5; Length 112;
Best Local Similarity 100.0%; Pred. No. 1.4e-104;
Matches 112; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 591 PVMNGKLNSSVLVPYGAESSESDSESKGLGKENGITGVSHSPGQDADEEATPHE 650
DB 1 PVMNGKLNSSVLVPYGAESSESDSESKGLGKENGITGVSHSPGQDADEEATPHE 60
QY 651 LOEPTLNGANSADSDPKENGLAPDGASCQGPALHSENFPAKANGLPCK 702
DB 61 LOEPTLNGANSADSDPKENGLAPDGASCQGPALHSENFPAKANGLPCK 112

RESULT 17
AAM18396
ID AAM18396 standard; protein; 66 AA.
AC AAM18396;
XX
XX
DT 12-OCT-2001 (first entry)
XX
DE Peptide #4830 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX Homo sapiens.
XX
XX WO200157277-A2.
PN
XX
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US000670.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX

(MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-488901/53.
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human cervical epithelial cells.
Claim 27; SEQ ID NO 2322; 487pp; English.
The present invention relates to human single exon nucleic acid probes (SENP: see AAI10068-AA12345). The present sequence is a peptide encoded by one such probe. The SENPs are derived from human Hela cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a sample derived from human cervical epithelial cells. By measuring gene expression, the probes are therefore useful in grading and/or staging of diseases of the cervix, notably cervical cancer. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 66 AA;
Query Match 8.5%; Score 66; DB 4; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.2e-58;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 82 LGDGIAPPQKVLFPSEKICLKWOOTHRVGAGLQNLGNTCFANALQCLTTPPLANTYMLS 141
DB 1 LGDGIAPPQKVLFPSEKICLKWOOTHRVGAGLQNLGNTCFANALQCLTTPPLANTYMLS 60
QY 142 HEHST 147
DB 61 HEHST 66
RESULT 18
ABB37427
ID ABB37427 standard; peptide; 66 AA.
AC ABB37427;
XX
XX 04-FEB-2002 (first entry)
DT
XX Peptide #4933 encoded by human foetal liver single exon probe.
DE Human; foetal liver; gene expression; single exon nucleic acid probe.
KW Homo sapiens.
XX
XX WO200157277-A2.
PN
XX 09-AUG-2001.
PD
XX 30-JAN-2001; 2001WO-US000669.
PF
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
Penn SG, Hanzel DK, Chen W, Rank DR;
WPI; 2001-483447/52.
Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver.
PT

XX
XX
XX
Claim 27; SEQ ID NO 30062; 639pp + Sequence Listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a peptide encoded by a single exon nucleic acid probe of the invention. Note: the sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
SQ Sequence 66 AA;

Query Match 8.5%; Score 66; DB 4; Length 66;
Best Local Similarity 100.0%; Pred.No. 3.2e-58;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 82 LGDGIAPPQKVLFPSEKICLKWOOTHRVGAGLQNLGNTCFANAALQCITVTPPIANTMLS 141
| | | | |
Db 1 LGDGIAPPQKVLFPSEKICLKWOOTHRVGAGLQNLGNTCFANAALQCITVTPPIANTMLS 60
| | | | |

QY 142 HEHSKT 147
| | | | |
Db 61 HEHSKT 66
| | | | |

RESULT 19
AAM30872
ID AAM30872 standard; protein; 66 AA.
XX
AC AAM30872;
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #4909 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
KW genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
PP
PD 09-AUG-2001.
PF 30-JAN-2001; 2001WQ-US000663.
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GH-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta.
PT
PS Claim 27; SEQ ID NO 31141; 654pp; English.
XX
XX The present invention relates to single exon nucleic acid probes (SENP: see AA131315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders

```

XX      SQ      Sequence 66 AA;
XX
XX      Query Match      8.5%;      Score 66;      DB 4;      Length 66;
XX      Best Local Similarity 100.0%;      Pred. No. 3.2e-58;
XX      Matches 66;      Conservative 0;      Mismatches 0;      Indels 0;      Gaps 0;
XX
XX      QY      82      LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTTPPIANTMLS 141
XX      DB      1      LGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTTPPIANTMLS 60
XX
XX      QY      142      HEHSKT 147
XX      DB      61      HEHSKT 66
XX
XX      RESULT 20
XX      ABB32185
XX      ID      ABB32185 standard; peptide; 66 AA.
XX
XX      AC      ABB32185;
XX
XX      DT      01-FEB-2002 (first entry)
XX
XX      DE      Peptide #4836 encoded by breast cell single exon nucleic acid probe.
XX
XX      KW      Human; microarray; single exon probe; gene expression; breast; disease;
XX      KW      cancer.
XX
XX      OS      Homo sapiens.
XX
XX      FN      WO200157271-A2.
XX
XX      PD      09-AUG-2001.
XX
XX      PF      30-JAN-2001; 2001WO-US0000662.
XX
XX      PR      04-FEB-2000; 2000US-0180312P.
XX      PR      26-MAY-2000; 2000US-0207456P.
XX      PR      30-JUN-2000; 2000US-00608408.
XX      PR      03-AUG-2000; 2000US-00632366.
XX      PR      21-SEP-2000; 2000US-0234687P.
XX      PR      27-SEP-2000; 2000US-0236359P.
XX      PR      04-OCT-2000; 2000GB-00024263.
XX
XX      (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX      Penn SG, Hanzel DK, Chen W, Rank DR,
XX
XX      WPI; 2001-496933/54.
XX
XX      New spatially-addressable set of single exon nucleic acid probes, useful
XX      for measuring gene expression in sample derived from human breast,
XX      comprises number of single exon nucleic acid probes.
XX
XX      Claim 27; SEQ ID NO 15153; 327pp + Sequence Listing; English.
XX
XX      The invention relates to a spatially-addressable set of single exon
XX      nucleic acid probes for measuring gene expression in a sample derived
XX      from human breast and BT 474 cells. The method involves contacting the
XX      probes with a collection of detectably labelled nucleic acids derived
XX      from mRNA of human breast, and then measuring the label bound to each
XX      probe of the microarray. The probes are useful for verifying the
XX      expression of regions of genomic DNA predicted to encode proteins. They
XX      are useful for gene discovery, and for determining predisposition and/or
XX      prognosing breast disease. Gene expression analysis is useful for
XX      assessing the toxicity of chemical agents on cells. The microarray of
XX      this invention presents a far greater diversity of probes for measuring
XX      gene expression, with far less bias than expressed sequence tag
XX      microarrays. The method is suitable for rapid production of functional
XX      information from genomic sequence. The present sequence is a peptide
XX      encoded by a single exon nucleic acid probe of the invention. Note: The
XX      sequence data for this patent did not form part of the printed

```

Db	1	LGDIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAAALQCLTYTPPLANYMLS	60
QY	142	HEHSKT 147	
Db	61	HEHSKT 66	
RESULT 22			
ABG52236			
ID	ABG52236	standard; peptide; 66 AA.	
AC	ABG52236;		
XX	25-FEB-2003	(first entry)	
XX	Human liver peptide, SEQ ID No 30884.		
DE	Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;		
KW	hypercholesterolaemia; coronary heart disease.		
OS	Homo sapiens.		
XX	W0200157273-A2.		
PN	09-AUG-2001.		
PD	30-JAN-2001; 2001WO-US000664.		
XX	04-FEB-2000; 2000US-0180312P.		
XX	26-MAY-2000; 2000US-0207456P.		
PR	30-JUN-2000; 2000US-00608408.		
PR	03-AUG-2000; 2000US-0063236P.		
PR	21-SEP-2000; 2000US-0234687P.		
PR	27-SEP-2000; 2000US-0236359P.		
PR	04-OCT-2000; 2000GB-00024263.		
XX	(MOLE-) MOLECULAR DYNAMICS INC.		
PA	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX	WPI; 2001-488898/53.		
XX	Human genome-derived single exon nucleic acid probes useful for analyzing		
PT	gene expression in human adult liver.		
XX	Claim 27; SEQ ID NO 30884; 658pp; English.		
XX	The invention relates to a single exon nucleic acid probe (SENP) (I) for		
CC	measuring human gene expression in a sample derived from human adult		
CC	liver, comprising one of 13109 defined nucleotide sequences given in the		
CC	specification (or complements/ fragments). The probe hybridises at high		
CC	stringency to a nucleic acid molecule expressed in the human adult liver.		
CC	(I) may be used for predicting, measuring and displaying gene expression		
CC	in samples derived from human adult liver. The genes identified may be		
CC	involved in genetic liver diseases such as cirrhosis,		
CC	hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is		
CC	associated with coronary heart disease. ABG47348-ABG59930 represent human		
CC	liver single exon encoded peptides of the invention. Note: The sequence		
CC	information for this patent does not appear in the printed specification		
CC	but was obtained in electronic format directly from WIPO at		
CC	ftp.wipo.int/pub/published_pct_sequences		
XX	Sequence 66 AA;		
XX	Query Match	8.5%; Score 66; DB 4; Length 66;	
XX	Best Local Similarity	100.0%; Pred. No. 3.2e-58;	
XX	Matches	66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	82	LGDIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAAALQCLTYTPPLANYMLS	141
Db	1	LGDIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAAALQCLTYTPPLANYMLS	60
QY	142	HEHSKT 147	

Db 61 HEHST 66

RESULT 23
AAM05997
ID AAM05997 standard; protein; 66 AA.
XX AC AAM05997;
XX DT 09-OCT-2001 (first entry)
XX DE Peptide #4679 encoded by probe for measuring breast gene expression.
XX KW Probe; human; breast disease; breast cancer; development disorder;
XX KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX OS Homo sapiens.
XX PN WO200157270-A2.
XX PD 09-AUG-2001.
XX PF 29-JAN-2001; 2001WO-US000661.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX Claim 27; SEQ ID NO 14737; 322pp; English.
XX The present invention relates to novel single exon nucleic acid probes
XX (see AAL00010-AA110067). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for measuring human gene expression in
XX a human breast sample, where the probe hybridises at high stringency to a
XX nucleic acid expressed in the human breast. The probes are useful for
XX predicting, diagnosing, grading, staging, monitoring and prognosing
XX diseases of the human breast, particularly those diseases with polygenic
XX aetiology. The diseases include: breast cancer, disorders of development,
XX inflammatory diseases of the breast, fibrocystic changes, proliferative
XX breast disease and non-carcinoma tumours. Note: The sequence data for
XX this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 66 AA;
XX
XX Query Match 8.5%; Score 66; DB 4; Length 66;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-58;
XX Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 82 LGDGTAPQKULFPSEKLCWKQOQTHRVGAGLQNLGNTCFANALQCLTYTPPLANYMLS 141
Db 1 LGDGTAPQKULFPSEKLCWKQOQTHRVGAGLQNLGNTCFANALQCLTYTPPLANYMLS 60
QY 142 HEHST 147
Db 61 HEHST 66

RESULT 24
ABG40214
ID ABG40214 standard; peptide; 66 AA.
XX AC ABG40214;
XX DT 19-AUG-2002 (first entry)
XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 29879.
XX KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
XX KW chronic obstructive pulmonary disease; interstitial lung disease;
XX KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
XX KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX KW primary ciliary dyskinesia; pulmonary hypertension;
XX KW hyaline membrane disease.
XX OS Homo sapiens.
XX PN WO200186003-A2.
XX PD 15-NOV-2001.
XX PF 30-JAN-2001; 2001WO-US000665.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2002-114183/15.
XX Spatially-addressable set of single exon nucleic acid probes, used to
XX measure gene expression in human lung samples.
XX Claim 27; SEQ ID NO 29879; 634pp; English.
XX The invention relates to a spatially-addressable set of single exon
XX nucleic acid probes for measuring gene expression in a sample derived
XX from human lung comprising single exon nucleic acid probes having one of
XX 12614 nucleic acid sequences mentioned in the specification, or their
XX complements or the 12387 open reading frames derived from the 12614
XX probes. Also included are a microarray comprising the novel set of probes
XX; the novel set of probes which hybridise at high stringency to a nucleic
XX acid expressed in the human lung; measuring gene expression in a sample
XX derived from human lung, comprising (a) contacting the array with a
XX collection of detectably labeled nucleic acids derived from human lung
XX mRNA, and (b) measuring the label detectably bound to each probe of the
XX array; identifying exons in a eukaryotic genome, comprising (a)
XX algorithmically predicting at least one exon from genomic sequences of
XX the eukaryote; and (b) detecting specific hybridisation of detectably
XX labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
XX having a fragment identical to the predicted exon, the probe is included
XX in the above mentioned microarray; assigning exons to a single gene,
XX comprising (a) identifying exons from genomic sequence by the method
XX above and (b) measuring the expression of each of the exons in several
XX tissues and/or cell types using hybridisation to a single exon
XX microarrays having a probe with the exon, where a common pattern of
XX expression of the exons in the tissues and/or cell types indicates that
XX the exons should be assigned to a single gene; a peptide comprising one
XX of 12011 sequences, mentioned in the specification, or encoded by the
XX probes/open reading frames (ORF). The probes are used for gene expression
XX analysis, and for identifying exons in a gene, particularly using human

Wed Aug 18 13:53:03 2004

CC lung derived mRNA and for the study of lung diseases such as asthma, lung
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
CC present sequence is a peptide/protein encoded by a single exon probe of
CC the invention. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 66 AA;

Query Match 8.5%; Score 66; DB 5; Length 66;
Best Local Similarity 100.0%; Pred. No. 3.2e-58;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 82 LGDGIAPPQKVPFSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 141
Db 1 LGDGIAPPQKVPFSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLS 60
QY 142 HEHSKT 147
Db 61 HEHSKT 66

RESULT 25
AAM84526
ID AAM84526 standard; protein; 133 AA.

XX AAM84526;
XX 07-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen SEQ ID NO:12119.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX Homo sapiens.

XX WO200157182-A2.
XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
XX 16-MAR-2000; 2000US-0189874P.
XX 17-MAR-2000; 2000US-0190076P.
XX 18-APR-2000; 2000US-0198123P.
XX 19-MAY-2000; 2000US-0205515P.
XX 07-JUN-2000; 2000US-0209467P.
XX 28-JUN-2000; 2000US-0214886P.
XX 30-JUN-2000; 2000US-0215135P.
XX 07-JUL-2000; 2000US-0216647P.
XX 07-JUL-2000; 2000US-0216880P.
XX 11-JUL-2000; 2000US-0217487P.
XX 11-JUL-2000; 2000US-0217496P.
XX 14-JUL-2000; 2000US-0218290P.
XX 26-JUL-2000; 2000US-0220963P.
XX 26-JUL-2000; 2000US-0220964P.
XX 14-AUG-2000; 2000US-0224518P.
XX 14-AUG-2000; 2000US-0224519P.
XX 14-AUG-2000; 2000US-0225213P.
XX 14-AUG-2000; 2000US-0225214P.
XX 14-AUG-2000; 2000US-0225266P.
XX 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-022679P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-023263P.
PR 14-SEP-2000; 2000US-023264P.
PR 14-SEP-2000; 2000US-023265P.
PR 21-SEP-2000; 2000US-0232223P.
PR 21-SEP-2000; 2000US-0232274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0245474P.
PR 08-NOV-2000; 2000US-0245475P.
PR 08-NOV-2000; 2000US-0245476P.
PR 08-NOV-2000; 2000US-0245477P.
PR 08-NOV-2000; 2000US-0245478P.
PR 08-NOV-2000; 2000US-0245523P.
PR 08-NOV-2000; 2000US-0245524P.
PR 08-NOV-2000; 2000US-0245525P.
PR 08-NOV-2000; 2000US-0245526P.

Query Match 5.8%; Score 45; DB 4; Length 133;

RESULT 27

ADC27049
ID ADC27049 standard; protein; 248 AA.
XX
AC
AC
AC
DT 18-DEC-2003 (first entry)
XX
XX
XX
DE Human deubiquitinating enzyme DUB8.10 SEQ ID NO:46.
XX
XX
KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
KW infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.
XX
XX
OS Homo sapiens.
XX
PN WO2003072724-A2.
XX
XX
XX 04-SEP-2003.
XX
XX 20-FEB-2003; 2003WO-US005338.
XX
XX
XX
XX 22-FEB-2002; 2002US-0358873P.
XX
XX 22-FEB-2002; 2002US-0358873P.
XX
XX 08-MAR-2002; 2002US-0363020P.
XX
XX 12-APR-2002; 2002GB-00008404.
XX
XX
XX (AVET) AVENTIS PHARM INC.
XX
XX
XX Hahn C, Liu H;
XX
XX
XX WPI; 2003-721759/68.
XX
XX N-PSDB; ADC27050.
XX
XX
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX
XX
XX Claim 2; SEQ ID NO 46; 201pp; English.
XX
XX
XX The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be used for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB8.10 from the present invention.
XX
SQ Sequence 248 AA;
Query Match 2.1%; Score 16; DB 7; Length 248;
Best Local Similarity 100.0%; Pred. No. 4.5e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 126 LQCLTYTPPLANYMLS 141
Db 95 LQCLTYTPPLANYMLS 110
RESULT 28
ID ADC27047
AC ADC27047 standard; protein; 271 AA.
XX
AC
AC
DT 18-DEC-2003 (first entry)
XX
XX
XX Human deubiquitinating enzyme DUB8.2 SEQ ID NO:44.
XX
XX

KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
KW infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.
XX
XX
XX Homo sapiens.
XX
PN WO2003072724-A2.
XX
XX
XX 04-SEP-2003.
XX
XX 20-FEB-2003; 2003WO-US005338.
XX
XX
XX
XX 22-FEB-2002; 2002US-0358873P.
XX
XX 22-FEB-2002; 2002US-0358873P.
XX
XX 08-MAR-2002; 2002US-0363020P.
XX
XX 12-APR-2002; 2002GB-00008404.
XX
XX
XX (AVET) AVENTIS PHARM INC.
XX
XX
XX Hahn C, Liu H;
XX
XX
XX WPI; 2003-721759/68.
XX
XX N-PSDB; ADC27046.
XX
XX
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX
XX
XX Claim 2; SEQ ID NO 44; 201pp; English.
XX
XX
XX The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be used for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB8.2 from the present invention.
XX
SQ Sequence 271 AA;
Query Match 2.1%; Score 16; DB 7; Length 271;
Best Local Similarity 100.0%; Pred. No. 4.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 126 LQCLTYTPPLANYMLS 141
Db 129 LQCLTYTPPLANYMLS 144
RESULT 29
ID ADA50458
AC ADA50458 standard; protein; 308 AA.
XX
AC
AC
DT 20-NOV-2003 (first entry)
XX
XX
XX Human protease SEQ ID NO:56.
XX
XX
XX enzyme; human; protease.
XX
XX
XX Homo sapiens.
XX
XX WO2003040393-A2.
XX
XX
XX 15-MAY-2003.
XX
XX
XX 04-NOV-2002; 2002WO-IB004615.
XX
XX

XX 06-NOV-2001; 2001US-0332633P.
 XX (DECO-) DECODE GENETICS EHP.
 PA Martinez RAM, Sigurdson GT;
 PI WPI; 2003-441582/41.
 XX N-PSDB; ADA50505.
 DR Novel isolated protease polypeptide and polynucleotide encoding the
 PT polypeptide useful for diagnosing and treating diseases or conditions
 PT associated with a protease.
 XX Claim 1; Page 66; 160pp; English.
 PS The invention relates to a novel isolated polypeptide comprising an amino
 CC acid sequence that has greater than 95 % identity to any one of 47 150-
 CC 350 residue protease polypeptide sequences, given in the specification.
 CC The nucleic acids, probes, primers, polypeptides and antibodies of the
 CC invention can be used in methods of diagnosis of a susceptibility to a
 CC disease or condition associated with a protease. The present sequence
 CC represents a protease of the invention.
 XX Sequence 308 AA;
 SQ
 Query Match 2.1%; Score 16; DB 6; Length 308;
 Best Local Similarity 100.0%; Pred. No. 5.5e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 Db 26 LQCLTYTPPLANYMLS 41
 RESULT 30
 AAEE19174
 ID AAEE19174 standard; protein; 314 AA.
 XX AAEE19174;
 AC AAEE19174;
 XX 21-MAY-2002 (first entry)
 DT Human protease, PRTS-11 protein.
 DE Human; protease; PRTS-11; enzyme; gastritis; cirrhosis; Crohn's disease;
 XX gastrointestinal disorder; autoimmune; inflammatory; cell proliferative;
 KW cardiovascular; developmental; epithelial; neurological; reproductive;
 KW AIDS; Acquired Immune Deficiency Syndrome; allergy; rheumatoid arthritis;
 KW anaemia; asthma; atherosclerosis; hypertension; myocardial infarction;
 KW hepatitis; cancer; psoriasis; Cushing's syndrome; hypothyroidism; eczema;
 KW epilepsy; Alzheimer's disease; Huntington's disease; Parkinson's disease;
 KW Pick's disease; infertility; vitiligo; drug screening; gene therapy;
 KW Chromosome 16.
 XX Homo sapiens.
 OS WO200208396-A2.
 PN 31-JAN-2002.
 PD 17-JUL-2001; 2001WO-US022397.
 XX 21-JUL-2000; 2000US-0220063P.
 PR 28-JUL-2000; 2000US-0221680P.
 PR 04-AUG-2000; 2000US-0223544P.
 PR 11-AUG-2000; 2000US-0224717P.
 PR 16-AUG-2000; 2000US-0225988P.
 PR 23-AUG-2000; 2000US-0227568P.
 XX (INCY-) INCYTE GENOMICS INC.
 PA Delegeane AM, Gandhi AR, Hafalia AJA, Lu DAM, Patterson C;
 XX

PI Tribouley CM, Das D, Kallick DA, Nguyen DB, Lee EA, Khan FA, Yue H;
 PI Au-Young J, Griffin JA, Policky JL, Ramkumar J, Yang J;
 PI Thangavelu K, Ding L, Kearney L, Baughn MR, Borowsky ML;
 PI Sanjanwala MS, Yao MG, Burford N, Walia NK, Lal P, Lee S, Todd S;
 PI Lo TP, Tang YT, Elliott VS, Azimzai Y, Lu Y;
 XX WPI; 2002-206082/26.
 DR N-PSDB; AAD30578.
 XX New human protease polypeptide, useful in diagnosis, prevention and
 PT treatment of gastrointestinal, cardiovascular, autoimmune/inflammatory,
 PT cell proliferative, developmental, epithelial and neurological disorders.
 XX Claim 1; Page 148; 182pp; English.
 PS The invention relates to an isolated human protease polypeptide (PRTS).
 CC PRTS protein and DNA are useful for diagnosing, treating and preventing
 CC gastrointestinal disorders (gastritis, cirrhosis, Crohn's disease),
 CC autoimmune/inflammatory disorders (AIDS, allergy, rheumatoid arthritis,
 CC anaemia, asthma), cardiovascular disorder (atherosclerosis, hypertension,
 CC myocardial infarction), cell proliferative disorders (hepatitis, cancer,
 CC psoriasis), developmental disorders (Cushing's syndrome, hypothyroidism),
 CC epithelial disorder (vitiligo, keloid, eczema), neurological disorders
 CC (epilepsy, Alzheimer's disease, Pick's disease, Huntington's disease,
 CC Parkinson's disease), and reproductive disorders (infertility). PRTS
 CC protein is useful in a number of drug screening techniques and to analyse
 CC the proteome of a tissue or cell type. PRTS DNA is useful for creating
 CC knockin humanised animals or transgenic animals to model human diseases,
 CC in somatic or germline gene therapy and in microarrays utilising fluids
 CC or tissues from patients to detect altered PKIN expression. The present
 CC sequence is human PRTS-11 protein
 XX Sequence 314 AA;
 SQ
 Query Match 2.1%; Score 16; DB 5; Length 314;
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 Db 95 LQCLTYTPPLANYMLS 110
 RESULT 31
 ADC27043
 ID ADC27043 standard; protein; 357 AA.
 XX ADC27043;
 AC ADC27043;
 XX 18-DEC-2003 (first entry)
 DT Human deubiquitinating enzyme DUB4.4 SEQ ID NO:40.
 DE human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX Homo sapiens.
 OS WO2003072724-A2.
 PN 04-SEP-2003.
 PD 20-FEB-2003; 2003WO-US005338.
 XX 22-FEB-2002; 2002US-0358873P.
 PR 22-FEB-2002; 2002US-0358875P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX (AVET) AVENTIS PHARM INC.
 PA Hahn C, Liu H;
 PI

DR WPI; 2003-721759/68.
DR N-PSDB; ADC27042.
XX
PT New isolated polynucleotide encoding human deubiquitinating proteases,
PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
PT also the encoded polypeptides.
XX
XX Claim 2; SEQ ID NO 40; 201pp; English.
XX
XX The present invention describes human deubiquitinating protease enzymes
CC (I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.4 from the present invention.
XX
XX Sequence 357 AA;
SQ

Query Match 2.1%; Score 16; DB 7; Length 357;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 LQCLTYTPPLANYMIS 141
Db 95 LQCLTYTPPLANYMIS 110

RESULT 32
ADC27045
ID ADC27045 standard; protein; 398 AA.
XX
AC ADC27045;
XX
XX 18-DEC-2003 (first entry)
XX
XX Human deubiquitinating enzyme DUB4.9 SEQ ID NO:42.
XX
XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
XX
XX Homo sapiens.
XX
XX WO2003072724-A2.
XX
XX 04-SEP-2003.
XX
XX 20-FEB-2003; 2003WO-US005338.
XX
XX 22-FEB-2002; 2002US-0358873P.
XX 22-FEB-2002; 2002US-0358873P.
XX 08-MAR-2002; 2002US-0363020P.
XX 12-APR-2002; 2002GB-00008404.
XX
XX (AVET) AVENTIS PHARM INC.
XX
XX Hahn C, Liu H;
XX
XX WPI; 2003-721759/68.
XX N-PSDB; ADC27044.
XX
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX
XX Claim 2; SEQ ID NO 42; 201pp; English.
XX

CC The present invention describes human deubiquitinating protease enzymes
CC (I), designated hDUB. The enzymes designated hDUB4 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present
CC sequence represents hDUB4.9 from the present invention.
XX
XX Sequence 398 AA;
SQ

Query Match 2.1%; Score 16; DB 7; Length 398;
Best Local Similarity 100.0%; Pred. No. 7e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 126 LQCLTYTPPLANYMIS 141
Db 95 LQCLTYTPPLANYMIS 110

RESULT 33
AAY77471
ID AAY77471 standard; protein; 530 AA.
XX
AC AAY77471;
XX
XX 05-JUN-2000 (first entry)
XX
XX Human deubiquitinating protein Dub11, SEQ ID NO:34.
XX
XX Human; deubiquitinating protein; Dub11; Dub12; immune disorder;
XX inflammation; allergy; immunosuppressant; antiarthritic; antirheumatoid;
XX antiinflammatory; dermatological; antithyroid.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Misc-difference 279 /label= Xaa
XX /note= 'Xaa is unknown'
XX
XX WO200001019-A2.
XX
XX 13-JAN-2000.
XX
XX 06-JUL-1998; 99WO-US012366.
XX
XX 06-JUL-1998; 98US-00110938.
XX 13-JUL-1998; 98US-00114466.
XX 23-JUL-1998; 98US-0093897P.
XX 12-AUG-1998; 98US-00132968.
XX 18-AUG-1998; 98US-00136214.
XX 11-SEP-1998; 98US-0093999P.
XX
XX (SCHE) SCHERING CORP.
XX
XX Bates EEM, Lebecque SE, Murphy EE, Mattson JD, Gorman DM;
XX Hedrick JA, Wang L, Zlotnik A, Murgolo N, Greene JR, Johnston JA;
XX Bazan JF, Mahony D, Lees EM;
XX
XX WPI; 2000-171015/15.
XX N-PSDB; AAZ92414.
XX
XX New isolated mammalian genes, used to develop products for treating e.g.
XX immune, inflammatory or allergic abnormalities, cancers or degenerative
XX conditions.
XX
XX Claim 43; Page 187-189; 218pp; English.
XX

XX The invention relates to a number of primate and/or rodent proteins, and
 CC the genes which encode them. The invention encompasses human dendritic
 CC cell prostaglandin transporter (DC-PGT); the TNF (tumour necrosis factor)
 CC receptor family-related proteins HPTFA84, HSLJD37R and RANKL; human CC
 CC chemokine HCC5; human deubiquitinating proteins Dub11 and Dub 12; human
 CC MD-1 and human and murine MD-2 proteins, which exhibit the properties of
 CC ligands for proteins comprising a leucine-rich motif (LRR); human cyclin
 CC E2; cDNAs encoding these proteins; and antibodies against these proteins.
 CC The proteins can be used for modulating the physiology or development of
 CC a cell. They can be used to mediate uptake of substrates (e.g.,
 CC prostaglandin-like molecules), to modulate or mediate cellular
 CC interactions (e.g., induce or prevent trafficking, proliferation, or
 CC differentiation of cells), or are intracellular proteins which are
 CC important in various cellular processes such as the deubiquitination of
 CC proteins or cell cycle regulation. The products can be used for treating
 CC medical conditions such as immune, inflammatory or allergic disorders, or
 CC abnormal cellular proliferation, for example, cancers or degenerative
 CC conditions. They can be used to modulate immune responses in disease
 CC states e.g., autoimmune disorders, including rheumatoid arthritis,
 CC systemic lupus erythematosus, Hashimoto's autoimmune thyroiditis, as well
 CC as acute and chronic inflammatory responses in which T cell activation,
 CC expansion, and/or immunological T cell memory play an important role.
 CC Sequences AAY77470-Y77473 represent human deubiquitinating proteins
 CC (Dub). AAY77470-Y77471 encode Dub11 proteins and AAY77472-Y77473 encode
 CC Dub12 proteins
 XX
 SQ Sequence 530 AA;

Query Match 2.1%; Score 16; DB 3; Length 530;
 Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTVTPPLANYMLS 141
 DB 95 LQCLTVTPPLANYMLS 110
 |||||

RESULT 34
 ASUG61836
 ID ASUG61836 standard; protein; 530 AA.

XX AC ASUG61836;

XX 15-AUG-2003 (first entry)

XX Updated human de-ubiquitination protein Dub11.

XX Human; enzyme; de-ubiquitination; Dub11; cell development;
 KW tissue culture development.

XX Homo sapiens.

XX Key Location/Qualifiers
 FT Misc-difference 279
 FT /label= Unknown
 FT /note= "Encoded by GTM"

PN US2003028005-A1.

PD 06-FEB-2003.

XX 11-AUG-1999; 99US-00372348.

XX 12-AUG-1998; 98US-0096328P.

XX (BAZA/) BAZAN J F.

XX Bazan JF;

XX WPI; 2003-466159/44.

XX N-P5DB; ACA62458.

PT New composition comprising HCC5, Dub11, Dub12, primate MD1 or MD2 or
 FT rodent MD2 polypeptide, useful for modulating physiology or development
 XX of a cell or tissue culture cells.

XX Claim 1; Page 8-10; 77pp; English.

XX The invention relates to a novel chemokine, de-ubiquitination or cell
 CC surface protein. The composition is useful for modulating physiology or
 CC development of a cell or tissue culture cells. The present sequence
 CC represents the amino acid sequence of the updated human de-ubiquitination
 CC protein Dub11

XX Sequence 530 AA;

Query Match 2.1%; Score 16; DB 6; Length 530;
 Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTVTPPLANYMLS 141
 DB 95 LQCLTVTPPLANYMLS 110
 |||||

RESULT 35

ABU10207
 ID ABU10207 standard; protein; 530 AA.

XX AC ABU10207;

XX 28-JUL-2003 (first entry)

XX Human ubiquitin carboxyl-terminal hydrolase 68999.

XX Human; enzyme; ubiquitin carboxyl-terminal hydrolase; tissue typing;
 KW cell proliferation disorder; cell differentiation disorder; cancer;
 KW hormonal disorder; neurological disorder; Parkinson's disease; rickets;
 KW Alzheimer's disease; viral infection; bacterial infection; diabetes;
 KW parasitic infection; bone metabolism disorder; osteoporosis; arrhythmia;
 KW immune disorder; multiple sclerosis; cardiovascular disorder; psoriasis;
 KW congestive heart failure; blood vessel disorder; hypertension; cachexia;
 KW Kawasaki syndrome; endothelial cell disorder; Grave's disease; obesity;
 KW liver disorder; hepatitis; Gaucher's disease; pain disorder;
 KW metabolic disorder.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Modified-site 22..24
 FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 23..26
 FT /note= "Casein kinase II phosphorylation site"

FT Modified-site 36..39
 FT /note= "Casein kinase II phosphorylation site"

FT Modified-site 71..73
 FT /note= "Protein kinase C phosphorylation site"

FT Domain 80..111
 FT /label= UCH1

FT Domain 81..96
 FT /note= "Ubiquitin carboxyl-terminal hydrolase-1 domain"

FT Domain 81..96
 FT /note= "Ubiquitin carboxyl-terminal hydrolase family 2
 signature 1 domain"

FT Modified-site 122..127

FT Binding-site 207..212
 FT /note= "N-myristoylation site"

FT Modified-site 212..217
 FT /note= "Cytochrome c family haem binding site"

FT Domain 264..279
 FT /note= "N-myristoylation site"

FT Modified-site 272..274
 FT /note= "Helix-loop-helix dimerisation domain"

FT Modified-site 281..284
 FT /note= "Protein kinase C phosphorylation site"

FT Modified-site 281..284
 FT /note= "cAMP and cGMP-dependent protein kinase"

FT phosphorylation site"
FT 288..293
FT /note= "N-myristoylation site"
FT Domain
FT 313..374
FT /label= UCH2
FT /note= "Ubiquitin carboxyl-terminal hydrolase-2 domain"
FT Domain
FT 317..335
FT /note= "Ubiquitin carboxyl-terminal hydrolase family 2
FT signature 2 domain"
FT 340..347
FT /note= "Tyrosine kinase phosphorylation site"
FT Modified-site
FT 376..379
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 395..400
FT /note= "N-myristoylation site"
FT Modified-site
FT 399..401
FT /note= "Protein kinase C phosphorylation site"
FT Modified-site
FT 401..404
FT /note= "cAMP and cGMP-dependent protein kinase
FT phosphorylation site"
FT 404..407
FT Modified-site
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 432..435
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 446..449
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 495..498
FT /note= "Casein kinase II phosphorylation site"
FT Modified-site
FT 503..508
FT /note= "N-myristoylation site"
FT Modified-site
FT 512..514
FT /note= "Protein kinase C phosphorylation site"
US2003022201-A1.
30-JAN-2003.
27-MAR-2002; 2002US-00107695.
27-MAR-2001; 2001US-0279184P.
(MILL-) MILLENNIUM PHARM INC.
Kapeller-Libermann R;
WPI; 2003-447421/42.
N-PSDB; ACA61698.
New 68999 nucleic acid molecule, useful for treating, preventing and/or
diagnosing 68999-mediated disorders, e.g. cancer, diabetes, arrhythmia,
pain or obesity, and in chromosome mapping, tissue typing and forensic
biology.
Claim 13; Page 40-41; 49pp; English.
The invention relates to an isolated 68999 nucleic acid molecule which
encodes a ubiquitin carboxyl-terminal hydrolase. The nucleic acid
molecules, polypeptides and antibodies are useful in screening assays,
predictive medicine (e.g. diagnostic assays, prognostic assays,
monitoring clinical trials and pharmacogenetics) and prophylactic and
therapeutic methods. The nucleic acids may also be used in chromosome
mapping, tissue typing and forensic biology and as surrogate markers. The
nucleic acid and polypeptide may be used for treating, preventing and/or
diagnosing 68999-mediated or -related disorders, such as cell
proliferation and/or differentiation disorders (e.g. cancers), hormonal
disorders, neurological disorders (e.g. Parkinson's or Alzheimer's
disease), infections (e.g. viral, bacterial or parasitic), bone
metabolism disorders (e.g. osteoporosis and rickets), immune disorders
(e.g. diabetes and multiple sclerosis), cardiovascular disorders (e.g.
arrhythmias congestive heart failure), blood vessel disorders (e.g.
hypertension and Kawasaki syndrome), endothelial cell disorder (e.g.
psoriasis and Grave's disease) liver disorders (e.g. hepatitis and
Gaucher's disease) and pain or metabolic disorders (e.g. obesity or

CC cachexia). The present sequence represents the amino acid sequence of
CC human ubiquitin carboxyl-terminal hydrolase 68999
XX Sequence 530 AA;
Query Match 2.1%; Score 16; DB 6; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 126 LQCLTTPPLANYMLS 141
DB 95 LQCLTTPPLANYMLS 110
RESULT 36
ADC27021
ID ADC27021 standard; protein; 530 AA.
XX AC ADC27021;
XX 18-DEC-2003 (first entry)
XX Human deubiquitinating enzyme DUB4.7 SEQ ID NO:18.
XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
XX Homo sapiens.
XX WO2003072724-A2.
XX 04-SEP-2003.
XX 20-FEB-2003; 2003WO-US045338.
XX 22-FEB-2002; 2002US-0358873P.
XX 22-FEB-2002; 2002US-0358875P.
XX 08-MAR-2002; 2002US-0361020P.
XX 12-APR-2002; 2002GB-00008404.
XX (AVET) AVENTIS PHARM INC.
XX Hahn C, Liu H;
XX WPI; 2003-721759/68.
XX N-PSDB; ADC27020.
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX Claim 2; SEQ ID NO 18; 201pp; English.
XX The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be used for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB4.7 from the present invention.
XX Sequence 530 AA;
Query Match 2.1%; Score 16; DB 7; Length 530;
Best Local Similarity 100.0%; Pred. No. 9.1e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
 |||
 Db 95 LQCLTYTPPLANYMLS 110

RESULT 37

ADC27033
 ID ADC27033 standard; protein; 530 AA.

AC ADC27033;

XX 18-DEC-2003 (first entry)

XX Human deubiquitinating enzyme DUB8.5 SEQ ID NO:30.

KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.

XX Homo sapiens.

OS WO2003072724-A2.

PN 04-SEP-2003.

PD 20-FEB-2003; 2003WO-US005338.

PF 22-FEB-2002; 2002US-0358873P.

PR 22-FEB-2002; 2002US-0358875P.

PR 08-MAR-2002; 2002US-0363020P.

PR 12-APR-2002; 2002GB-00008404.

XX (AVET) AVENTIS PHARM INC.

PA Hahn C, Liu H;

PI WPI; 2003-721759/68.

XX N-PSDB; ADC27032.

XX New isolated polynucleotide encoding human deubiquitinating proteases,

PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,

PT also the encoded polypeptides.

XX Claim 2; SEQ ID NO 30; 201pp; English.

XX The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB8.5 from the present invention.

XX Sequence 530 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141

Db 95 LQCLTYTPPLANYMLS 110

RESULT 38

ADC27019
 ID ADC27019 standard; protein; 530 AA.

XX

AC ADC27019;

XX 18-DEC-2003 (first entry)

DE Human deubiquitinating enzyme DUB4.6 SEQ ID NO:16.

XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.

XX Homo sapiens.

XX WO2003072724-A2.

PN 04-SEP-2003.

PD 20-FEB-2003; 2003WO-US005338.

PF 22-FEB-2002; 2002US-0358873P.

PR 22-FEB-2002; 2002US-0358875P.

PR 08-MAR-2002; 2002US-0363020P.

PR 12-APR-2002; 2002GB-00008404.

XX (AVET) AVENTIS PHARM INC.

PA Hahn C, Liu H;

PI WPI; 2003-721759/68.

XX N-PSDB; ADC27018.

XX New isolated polynucleotide encoding human deubiquitinating proteases,

PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,

PT also the encoded polypeptides.

XX Claim 2; SEQ ID NO 16; 201pp; English.

XX The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.6 from the present invention.

XX Sequence 530 AA;

Query Match

Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141

Db 95 LQCLTYTPPLANYMLS 110

RESULT 39

ADC27037
 ID ADC27037 standard; protein; 530 AA.

XX ADC27037;

XX 18-DEC-2003 (first entry)

XX Human deubiquitinating enzyme DUB8.7 SEQ ID NO:34.

XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.

OS Homo sapiens.
 XX WO2003072724-A2.
 XX PD 04-SEP-2003.
 XX PF 20-FEB-2003; 2003WO-US005338.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 08-MAR-2002; 2002US-0363020P.
 XX PR 12-APR-2002; 2002GB-00008404.
 XX PA (AVET) AVENTIS PHARM INC.
 XX PI Hahn C, Liu H;
 XX WPI; 2003-721759/68.
 XX N-PSDB; ADC27036.
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 XX also the encoded polypeptides.
 XX Claim 2; SEQ ID NO 34; 201pp; English.
 XX The present invention describes human deubiquitinating protease enzymes
 XX (I), designated hDUB. The enzymes designated hDUB4 are located on
 XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 XX 8. Also described is a method for reducing inflammation, modulating
 XX autoimmune diseases and modulating immune reactions during infection by
 XX administering a compound (A) that inhibits (I) or alters regulation of
 XX the transcription of nucleic acid sequences encoding (I). (I) has
 XX antiinflammatory and immunosuppressive activities, and can be used in the
 XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 XX proliferation. (I) can be used for identifying their respective
 XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
 XX autoimmune diseases or immune reactions during infections. The present
 XX sequence represents hDUB8.7 from the present invention.
 XX Sequence 530 AA;
 XX Query Match 2.1%; Score 16; DB 7; Length 530;
 XX Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 DB 95 LQCLTYTPPLANYMLS 110
 RESULT 40
 ADC27015
 ID ADC27015 standard; protein; 530 AA.
 AC ADC27015;
 XX 18-DEC-2003 (first entry)
 XX Human deubiquitinating enzyme DUB4.3 SEQ ID NO:12.
 XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX Homo sapiens.
 XX WO2003072724-A2.
 XX PD 04-SEP-2003.
 XX PF 20-FEB-2003; 2003WO-US005338.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 08-MAR-2002; 2002US-0363020P.
 XX PR 12-APR-2002; 2002GB-00008404.
 XX PA (AVET) AVENTIS PHARM INC.
 XX PI Hahn C, Liu H;
 XX WPI; 2003-721759/68.
 XX N-PSDB; ADC27036.
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 XX also the encoded polypeptides.
 XX Claim 2; SEQ ID NO 34; 201pp; English.
 XX The present invention describes human deubiquitinating protease enzymes
 XX (I), designated hDUB. The enzymes designated hDUB4 are located on
 XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 XX 8. Also described is a method for reducing inflammation, modulating
 XX autoimmune diseases and modulating immune reactions during infection by
 XX administering a compound (A) that inhibits (I) or alters regulation of
 XX the transcription of nucleic acid sequences encoding (I). (I) has
 XX antiinflammatory and immunosuppressive activities, and can be used in the
 XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 XX proliferation. (I) can be used for identifying their respective
 XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
 XX autoimmune diseases or immune reactions during infections. The present
 XX sequence represents hDUB8.7 from the present invention.
 XX Sequence 530 AA;
 XX Query Match 2.1%; Score 16; DB 7; Length 530;
 XX Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 DB 95 LQCLTYTPPLANYMLS 110
 RESULT 40
 ADC27015
 ID ADC27015 standard; protein; 530 AA.
 AC ADC27015;
 XX 18-DEC-2003 (first entry)
 XX Human deubiquitinating enzyme DUB4.3 SEQ ID NO:12.
 XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX Homo sapiens.
 XX WO2003072724-A2.
 XX PD 04-SEP-2003.
 XX PF 20-FEB-2003; 2003WO-US005338.
 XX PR 22-FEB-2002; 2002US-0358873P.

PR 22-FEB-2002; 2002US-0358873P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX (AVET) AVENTIS PHARM INC.
 XX Hahn C, Liu H;
 XX WPI; 2003-721759/68.
 XX N-PSDB; ADC27014.
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 XX also the encoded polypeptides.
 XX Claim 2; SEQ ID NO 12; 201pp; English.
 XX The present invention describes human deubiquitinating protease enzymes
 XX (I), designated hDUB. The enzymes designated hDUB4 are located on
 XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 XX 8. Also described is a method for reducing inflammation, modulating
 XX autoimmune diseases and modulating immune reactions during infection by
 XX administering a compound (A) that inhibits (I) or alters regulation of
 XX the transcription of nucleic acid sequences encoding (I). (I) has
 XX antiinflammatory and immunosuppressive activities, and can be used in the
 XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 XX proliferation. (I) can be used for identifying their respective
 XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
 XX autoimmune diseases or immune reactions during infections. The present
 XX sequence represents hDUB4.3 from the present invention.
 XX Sequence 530 AA;
 XX Query Match 2.1%; Score 16; DB 7; Length 530;
 XX Best Local Similarity 100.0%; Pred. No. 9.1e-07;
 XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 LQCLTYTPPLANYMLS 141
 DB 95 LQCLTYTPPLANYMLS 110
 RESULT 41
 ADC27023
 ID ADC27023 standard; protein; 565 AA.
 XX ADC27023;
 XX 18-DEC-2003 (first entry)
 XX Human deubiquitinating enzyme DUB4.8 SEQ ID NO:20.
 XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 XX infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX Homo sapiens.
 XX WO2003072724-A2.
 XX PD 04-SEP-2003.
 XX PF 20-FEB-2003; 2003WO-US005338.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 22-FEB-2002; 2002US-0358873P.
 XX PR 08-MAR-2002; 2002US-0363020P.
 XX PR 12-APR-2002; 2002GB-00008404.
 XX PA (AVET) AVENTIS PHARM INC.
 XX PI Hahn C, Liu H;
 XX WPI; 2003-721759/68.

DR N-PSDB; ADC27022.
 XX New isolated polynucleotide encoding human deubiquitinating proteases,
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.
 XX
 PS Claim 2; SEQ ID NO 20; 201pp; English.
 XX
 CC The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on
 CC chromosome 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.8 from the present invention.
 XX
 SQ Sequence 565 AA;

Query Match 2.1%; Score 16; DB 7; Length 565;
 Best Local Similarity 100.0%; Pred. No. 9.6e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
 |||||
 Db 130 LQCLTYTPPLANYMLS 145

RESULT 42

ADC27017
 ID ADC27017 standard; protein; 574 AA.

XX AC ADC27017;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.5 SEQ ID NO:14.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX Homo sapiens.
 OS
 XX WO2003072724-A2.
 PN
 PD 04-SEP-2003.
 XX
 PF 20-FEB-2003; 2003WO-US005338.
 XX
 PR 22-FEB-2002; 2002US-0358873P.
 PR 22-FEB-2002; 2002US-0358875P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX
 PA (AVET) AVENTIS PHARM INC.
 XX
 PI Hahn C, Liu H;
 XX
 DR WPI; 2003-721759/68.
 DR N-PSDB; ADC27016.
 XX
 PT New isolated polynucleotide encoding human deubiquitinating proteases,
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.
 XX
 PS Claim 2; SEQ ID NO 14; 201pp; English.
 XX
 CC The present invention describes human deubiquitinating protease enzymes

CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.5 from the present invention.
 XX
 SQ Sequence 574 AA;

Query Match 2.1%; Score 16; DB 7; Length 574;
 Best Local Similarity 100.0%; Pred. No. 9.8e-07;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 LQCLTYTPPLANYMLS 141
 |||||
 Db 130 LQCLTYTPPLANYMLS 145

RESULT 43

ADC27027
 ID ADC27027 standard; protein; 633 AA.

XX AC ADC27027;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Human deubiquitinating enzyme DUB4.11 SEQ ID NO:24.
 XX
 KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.
 XX Homo sapiens.
 OS
 XX WO2003072724-A2.
 PN
 PD 04-SEP-2003.
 XX
 PF 20-FEB-2003; 2003WO-US005338.
 XX
 PR 22-FEB-2002; 2002US-0358873P.
 PR 22-FEB-2002; 2002US-0358875P.
 PR 08-MAR-2002; 2002US-0363020P.
 PR 12-APR-2002; 2002GB-00008404.
 XX
 PA (AVET) AVENTIS PHARM INC.
 XX
 PI Hahn C, Liu H;
 XX
 DR WPI; 2003-721759/68.
 DR N-PSDB; ADC27026.

PT New isolated polynucleotide encoding human deubiquitinating proteases,
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.
 XX
 PS Claim 2; SEQ ID NO 24; 201pp; English.

CC The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on
 CC chromosome 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective

CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.11 from the present invention.

XX Sequence 633 AA;

Query Match 2.1%; Score 16; DB 7; Length 633;
 Best Local Similarity 100.0%; Pred. No. 1.7e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 LQCLTYTPPLANYMLS 141
 |||||
 Db 198 LQCLTYTPPLANYMLS 213
 |||||

RESULT 44

AD27011
 ID ADC27011 standard; protein; 1021 AA.

XX AC ADC27011;

DT 18-DEC-2003 (first entry)

DE Human deubiquitinating enzyme DUB4.2a SEQ ID NO:8.

XX human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
 KW infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.

XX OS Homo sapiens.

XX PN WO2003072724-A2.

XX PD 04-SEP-2003.

XX PF 20-FEB-2003; 2003WO-US005338.

XX PR 22-FEB-2002; 2002US-0358873P.

XX PR 22-FEB-2002; 2002US-0358875P.

XX PR 08-MAR-2002; 2002US-0363020P.

XX PR 12-APR-2002; 2002GB-00008404.

XX PA (AVET) AVENTIS PHARM INC.

XX PI Hahn C, Liu H;

XX DR WPI; 2003-721759/68.

XX DR N-PSDB; ADC27010.

XX PT New isolated polynucleotide encoding human deubiquitinating proteases,
 PT useful for identifying e.g. antiinflammatory or immunosuppressive agents,
 PT also the encoded polypeptides.

XX PS Claim 2; SEQ ID NO 8; 201pp; English.

XX CC The present invention describes human deubiquitinating protease enzymes
 CC (I), designated hDUB. The enzymes designated hDUB4 are located on
 CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
 CC 8. Also described is a method for reducing inflammation, modulating
 CC autoimmune diseases and modulating immune reactions during infections by
 CC administering a compound (A) that inhibits (I) or alters regulation of
 CC the transcription of nucleic acid sequences encoding (I). (I) has
 CC antiinflammatory and immunosuppressive activities, and can be used in the
 CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
 CC proliferation. (I) can be used for identifying their respective
 CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
 CC autoimmune diseases or immune reactions during infections. The present
 CC sequence represents hDUB4.2a from the present invention.

XX SQ Sequence 1021 AA;

Query Match 2.1%; Score 16; DB 7; Length 1021;
 Best Local Similarity 100.0%; Pred. No. 1.7e-06;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 126 LQCLTYTPPLANYMLS 141
 |||||
 Db 11 LQCLTYTPPLANYMLS 26
 |||||

RESULT 45

ADC31203
 ID ADC31203 standard; protein; 497 AA.

XX AC ADC31203;

DT 18-DEC-2003 (first entry)

DE Human novel polypeptide sequence, SEQ ID NO:1285.

XX Human; diagnostic; drug screening; forensics; gene mapping;
 KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
 KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
 KW ulcers; osteoporosis; autoimmune disease; cancer;
 KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
 KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
 KW anticulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
 KW gene therapy; chromosome 9p23.

XX OS Homo sapiens.

XX PN WO2003029271-A2.

XX PD 10-APR-2003.

XX PF 24-SEP-2002; 2002WO-US030474.

XX PR 24-SEP-2001; 2001US-0324631P.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;

XX PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;

XX PI Haley-Vicente D, Drmanic RT;

XX DR WPI; 2003-371981/35.

XX DR N-PSDB; ADC30232.

XX PT New polynucleotide and polypeptide useful for diagnosing, preventing or
 PT treating conditions such as neurodegenerative diseases, anemias, platelet
 PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
 PT cancer.

XX PS Claim 20; SEQ ID NO 1285; 1185pp; English.

XX CC The invention relates to 971 novel human cDNA sequences (ADC29919-
 CC ADC30889) and the polypeptides they encode (ADC30890-ADC31860). The
 CC invention also relates to nucleic acid sequences over 99% identical with
 CC the novel human cDNAs. The invention additionally encompasses expression
 CC vectors and host cells comprising a nucleic acid of the invention; the
 CC recombinant production of a polypeptide of the invention; an antibody
 CC against a polypeptide of the invention; a method of detecting
 CC polynucleotides or polypeptides of the invention; and methods of
 CC identifying a compound which binds to a polypeptide of the invention. The
 CC invention further discloses methods of preventing, treating or
 CC ameliorating a medical condition; kits comprising polynucleotide probes
 CC and/or monoclonal antibodies for carrying out the methods of the
 CC invention; methods for the identification of compounds that modulate the
 CC expression or activity of the polynucleotide and/or polypeptide; and 767
 CC contig sequences corresponding to the cDNA sequences of the invention
 CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
 CC -ADC33394). The nucleic acids and polypeptides of the invention are
 CC useful in diagnostics, drug screening, forensics, gene mapping, in the
 CC identification of mutations responsible for genetic disorders or other
 CC traits, for assessing biodiversity, and in producing many other types of
 CC data and products dependent on DNA and amino acid sequences. They are
 CC also used for treating diseases such as Parkinson's disease, Alzheimer's

CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers. And in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human polypeptide sequence of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX
SQ Sequence 497 AA;

Query Match 1.9%; Score 15; DB 7; Length 497;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 QCLTYTPPLANYMLS 141
Db 96 QCLTYTPPLANYMLS 110
|||||

RESULT 46
ADC27031
ID ADC27031 standard; protein; 497 AA.

XX AC ADC27031;
XX
XX
DT 18-DEC-2003 (first entry)
DE
DE Human deubiquitinating enzyme DUB8.3 SEQ ID NO:28.

XX KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease;
XX infection; antiinflammatory; immunosuppressive; hDUB8; chromosome 8.

XX OS Homo sapiens.
XX
XX
PN WO2003072724-A2.
XX
XX
PD 04-SEP-2003.
XX
XX
PF 20-FEB-2003; 2003WO-US005338.
XX
XX
PR 22-FEB-2002; 2002US-0358873P.
PR 22-FEB-2002; 2002US-0358875P.
PR 08-MAR-2002; 2002US-0363020P.
PR 12-APR-2002; 2002GB-00008404.

XX PA (AVET) AVENTIS PHARM INC.

XX PI Hahn C, Liu H;

XX DR WPI; 2003-721759/68.
XX DR N-PSDB; ADC27030.

XX PT New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX PT also the encoded polypeptides.

XX PS Claim 2; SEQ ID NO 28; 201pp; English.

XX CC The present invention describes human deubiquitinating protease enzymes
CC (1), designated hDUB. The enzymes designated hDUB8 are located on
CC chromosome 4, and the enzymes designated hDUB8 are located on chromosome
CC 8. Also described is a method for reducing inflammation, modulating
CC autoimmune diseases and modulating immune reactions during infections by
CC administering a compound (A) that inhibits (I) or alters regulation of
CC the transcription of nucleic acid sequences encoding (I). (I) has
CC antiinflammatory and immunosuppressive activities, and can be used in the
CC regulation of cytokine/cytokine receptor signaling involved in lymphocyte
CC proliferation. (I) can be used for identifying their respective
CC inhibitors (A). (A) are useful for reducing inflammation and to modulate
CC autoimmune diseases or immune reactions during infections. The present

CC sequence represents hDUB8.3 from the present invention.

XX Sequence 497 AA;

Query Match 1.9%; Score 15; DB 7; Length 497;
Best Local Similarity 100.0%; Pred. No. 8.8e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 127 QCLTYTPPLANYMLS 141
Db 96 QCLTYTPPLANYMLS 110
|||||

RESULT 47
AAAY77472
ID AAAY77472 standard; protein; 245 AA.

XX AC AAAY77472;
XX
XX
DT 05-JUN-2000 (first entry)

XX DE Human deubiquitinating protein Dub12.

XX KW Human; deubiquitinating protein; Dub11; Dub12; immune disorder;
XX inflammation; allergy; immunosuppressant; antiarthritic; antirheumatoid;
XX antiinflammatory; dermatological; antithyroid.

XX OS Homo sapiens.

XX PN WO200001817-A2.

XX PD 13-JAN-2000.

XX PF 06-JUL-1999; 99WO-US012366.

XX PR 06-JUL-1998; 98US-00110938.
XX PR 13-JUL-1998; 98US-00114466.
XX PR 23-JUL-1998; 98US-0093897P.
XX PR 12-AUG-1998; 98US-00132968.
XX PR 18-AUG-1998; 98US-00136214.
XX PR 11-SEP-1998; 98US-0099999P.

XX PA (SCHE) SCHERING CORP.

XX PI Bates EEM, Lebecque SJE, Murphy BE, Mattson JD, Gorman DM;

XX PI Hedrick JA, Wang L, Zlotnik A, Murgolo NJ, Greene JR, Johnston JA;
XX PI Bazan JF, Mahony D, Lees EM;

XX DR WPI; 2000-171015/15.
XX DR N-PSDB; AAZ92415.

XX PT New isolated mammalian genes, used to develop products for treating e.g.
XX immune, inflammatory or allergic abnormalities, cancers or degenerative
XX PT conditions.

XX PS Claim 43; Page 191-192; 218pp; English.

XX CC The invention relates to a number of primate and/or rodent proteins, and
XX the genes which encode them. The invention encompasses human dendritic
XX cell prostaglandin transporter (DC-PGT); the TNF (tumour necrosis factor)
XX receptor family-related proteins HPTA84, HSLJD37R and RANKL; human CC
XX chemokine HCC5; human deubiquitinating proteins Dub1 and Dub 12; human
XX MD-1 and human and murine MD-2 proteins, which exhibit the properties of
XX ligands for proteins comprising a leucine-rich motif (LRR); human cyclin
XX E2; cDNAs encoding these proteins; and antibodies against these proteins.
XX The proteins can be used for modulating the physiology or development of
XX a cell. They can be used to mediate uptake of substrates (e.g.,
XX prostaglandin-like molecules), to modulate or mediate cellular
XX interactions (e.g., induce or prevent trafficking, proliferation, or
XX differentiation of cells), or are intracellular proteins which are
XX important in various cellular processes such as the deubiquitination of
XX proteins or cell cycle regulation. The products can be used for treating
XX medical conditions such as immune, inflammatory or allergic disorders, or

CC abnormal cellular proliferation, for example, cancers or degenerative
 CC conditions. They can be used to modulate immune responses in disease
 CC states e.g., autoimmune disorders, including rheumatoid arthritis,
 CC systemic lupus erythematosus, Hashimoto's autoimmune thyroiditis, as well
 CC as acute and chronic inflammatory responses in which T cell activation,
 CC expansion, and/or immunological T cell memory play an important role.
 CC Sequences AAY77470-Y77473 represent human deubiquitinating proteins
 CC (Dub). AAY77470-Y77471 encode Dub1 proteins and AAY77472-Y77473 encode
 CC Dub12 proteins
 XX Sequence 245 AA;
 SQ Query Match 1.8%; Score 14; DB 3; Length 245;
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 310 SNVLTSLKRFANF 323
 Db ||||||||||||
 96 SNVLTSLKRFANF 109
 RESULT 48
 ABU61837
 ID ABU61837 standard; protein; 245 AA.
 XX AC ABU61837;
 XX DT 15-AUG-2003 (first entry)
 XX DE Human de-ubiquitination protein Dub12.
 XX KW Human; enzyme; de-ubiquitination; Dub12; cell development;
 XX KW tissue culture development.
 XX OS Homo sapiens.
 XX PH Key Location/Qualifiers
 FT Misc-difference 66
 FT /label= Unknown
 FT /note= "Encoded by GMA"
 FT Misc-difference 197
 FT /label= Unknown
 FT /note= "Encoded by TNA"
 FT Misc-difference 222
 FT /label= Unknown
 FT /note= "Encoded by NNN"
 FT Misc-difference 236
 FT /label= Unknown
 FT /note= "Encoded by AGN"
 FT Misc-difference 239
 FT /label= Unknown
 FT /note= "Encoded by NTC"
 FT US2003028005-A1.
 XX PD 06-FEB-2003.
 XX PF 11-AUG-1999; 99US-00372348.
 XX PR 12-AUG-1998; 98US-0096328P.
 XX PA (BAZA/) BAZAN J F.
 XX PI Bazan JF;
 XX DR WPI; 2003-466159/44.
 XX DR N-PSDB; ACA62459.
 XX New composition comprising HCC5, Dub11, Dub12, primate MD1 or MD2 or
 XX rodent MD2 polypeptide, useful for modulating physiology or development
 XX of a cell or tissue culture cells.
 XX Claim 1; Page 10-11; 77pp; English.

XX The invention relates to a novel chemokine, de-ubiquitination or cell
 CC surface protein. The composition is useful for modulating physiology or
 CC development of a cell or tissue culture cells. The present sequence
 CC represents the amino acid sequence of the human de-ubiquitination protein
 CC Dub12
 XX Sequence 245 AA;
 SQ Query Match 1.8%; Score 14; DB 6; Length 245;
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 310 SNVLTSLKRFANF 323
 Db ||||||||||||
 96 SNVLTSLKRFANF 109
 RESULT 49
 AAB95208
 ID AAB95208 standard; protein; 548 AA.
 XX AC AAB95208;
 XX DT 26-JUN-2001 (first entry)
 XX DE Human protein sequence SEQ ID NO:17313.
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX OS Homo sapiens.
 XX PN EP1074617-A2.
 XX PD 07-FEB-2001.
 XX PF 28-JUL-2000; 2000EP-00116126.
 XX PR 29-JUL-1999; 99JP-00248036.
 XX PR 27-AUG-1999; 99JP-00330253.
 XX PR 11-JAN-2000; 2000JP-00118776.
 XX PR 02-MAY-2000; 2000JP-00133767.
 XX PR 09-JUN-2000; 2000JP-00241899.
 XX PA (HELI-) HELIX RES INST.
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WIPI; 2001-318749/34.
 XX PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 XX length cDNAs defined in the specification, and for the detection and/or
 XX diagnosis of the abnormality of the proteins encoded by the full-length
 XX cDNAs.
 XX Claim 8; SEQ ID NO 17313; 2537pp + Sequence Listing; English.
 XX The present invention describes primer sets for synthesizing 5602 full-
 XX length cDNAs defined in the specification. Where a primer set comprises:
 XX (a) an oligo-dr primer and an oligonucleotide complementary to the
 XX complementary strand of a polynucleotide which comprises one of the 5602
 XX nucleotide sequences defined in the specification, where the
 XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 XX of an oligonucleotide comprising a sequence complementary to the
 XX complementary strand of a polynucleotide which comprises a 5'-end
 XX sequence and an oligonucleotide comprising a sequence complementary to a
 XX polynucleotide which comprises a 3'-end sequence, where the
 XX oligonucleotide comprises at least 15 nucleotides and the combination of
 XX the 5'-end sequence/3'-end sequence is selected from those defined in the
 XX specification. The primer sets can be used in antisense therapy and in
 XX gene therapy. The primers are useful for synthesizing polynucleotides,
 XX particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 548 AA;

Query Match 1.8%; Score 14; DB 4; Length 548;
 Best Local Similarity 100.0%; Pred. No. 9.9e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 310 SNVLTLSLKRFPANF 323
 |||||
 Db 321 SNVLTLSLKRFPANF 334

RESULT 50
 AAB95175
 ID AAB95175 standard; protein; 726 AA.
 XX
 AC AAB95175;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17235.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.

XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.

XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

XX
 PS Claim 8; SEQ ID NO 17235; 2537pp + Sequence Listing; English.

XX
 CC The present invention describes primer sets for synthesising 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesising polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention
 XX
 SQ Sequence 726 AA;

Query Match 1.8%; Score 14; DB 4; Length 726;
 Best Local Similarity 100.0%; Pred. No. 0.00013;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 310 SNVLTLSLKRFPANF 323
 |||||
 Db 21 SNVLTLSLKRFPANF 34

Search completed: August 10, 2004, 20:15:19
 Job time : 61 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:54:23 ; Search time 58 Seconds
(without alignments)
3775.419 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIVDKASESSDPAYQNQP.....AESLEPDAASLFFPSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A. Geneseq_29Jan04:*

1: geneseqp1980s:*

2: geneseqp1990s:*

3: geneseqp2000s:*

4: geneseqp2001s:*

5: geneseqp2002s:*

6: geneseqp2003as:*

7: geneseqp2003bs:*

8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4070	100.0	775	4 AAB74671	Aab74671 Human pro
2	4028	99.0	1251	5 AAU98889	Aau98889 Human pro
3	4022	98.8	1197	4 AAB95146	Aab95146 Human pro
4	4021	98.8	1270	5 ABP69291	Abp69291 Human pol
5	3999	98.3	762	4 AAB82177	Abp82177 Human ubi
6	3999	98.3	762	5 ABU08951	Abu08951 Human ubi
7	2908	71.4	558	5 ABB06118	Abb06118 Human NS
8	1423.5	35.0	1087	7 ADE31066	Ade31066 Human dia
9	1423.5	35.0	1123	4 AAY71889	Aay71889 Human ubi
10	1422.5	35.0	954	4 AAB93200	Aab93200 Human pro
11	1375	33.8	548	4 AAB95208	Aab95208 Human pro
12	1150.5	28.3	565	7 ADC27023	Adc27023 Human pro
13	1148	28.2	530	6 ABU10207	Abu10207 Human ubi
14	1142.5	28.1	530	7 ADC27019	Adc27019 Human ubi
15	1140.5	28.0	530	7 ADC27015	Adc27015 Human pro
16	1139.5	28.0	574	7 ADC27017	Adc27017 Human pro
17	1135.5	27.9	530	7 ADC27021	Adc27021 Human pro
18	1132.5	27.8	715	7 ADC27025	Adc27025 Human pro
19	1128.5	27.7	1021	7 ADC27011	Adc27011 Human pro
20	1128	27.7	530	4 AAG64049	Aag64049 Human pro
21	1127	27.7	494	2 AAW30711	Aaw30711 Human ubi
22	1126.5	27.7	633	7 ADC27027	Adc27027 Human pro
23	1121.5	27.6	1016	7 ADC27007	Adc27007 Human pro
24	1118.5	27.5	530	7 ADC27013	Adc27013 Human pro
25	1114	27.4	398	7 ADC27045	Adc27045 Human pro

26	1113.5	27.4	530	3 AAY77471	Aay77471 Human deu
27	1113.5	27.4	530	6 ABU61836	Abu61836 Updated h
28	1111.5	27.3	530	7 ADC27009	Adc27009 Human deu
29	1096	26.9	530	7 ADC27037	Adc27037 Human deu
30	1073	26.4	530	7 ADC27029	Adc27029 Human deu
31	1052	25.8	530	6 ABR43250	Abr43250 Human PMM
32	1052	25.8	530	7 ADC27035	Adc27035 Human deu
33	1049.5	25.8	530	7 ADC27039	Adc27039 Human deu
34	1037	25.5	357	7 ADC27043	Adc27043 Human deu
35	1036	25.5	308	6 ADA50458	Ada50458 Human pro
36	1030	25.3	530	7 ADC27033	Adc27033 Human pro
37	999	24.5	308	6 ADA50460	Ada50460 Human pro
38	980.5	24.1	497	7 ADC31203	Adc31203 Human nov
39	980.5	24.1	497	7 ADC27031	Adc27031 Human deu
40	914	22.5	748	3 AAY77473	Aay77473 Human deu
41	914	22.5	748	6 ABU61838	Abu61838 Updated h
42	889.5	21.9	545	2 AAW30708	Aaw30708 Murine ub
43	888	21.8	545	3 AAY77475	Aay77475 Primate p
44	888	21.8	545	6 ABU61840	Abu61840 Human de-
45	846.5	20.8	521	2 AAW30710	Aaw30710 Murine ub

ALIGNMENTS

RESULT 1
AAB74671
ID AAB74671 standard; protein; 775 AA.
XX
AC AAB74671;
XX
AC AAB74671;
XX
DT 12-JUN-2001 (first entry)
XX
DE Human protease and protease inhibitor PPIM-4.
XX
KW Human; protease; protease inhibitor; protease and protease inhibitor;
KW PPIM; identification; diagnosis; anti-human immunodeficiency virus; HIV;
KW antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
KW antithyroid; immunosuppressive; nephrotropic; antitumor; thyromimetic;
KW cytostatic; antibacterial; fungicide; protozoicide; hepatotropic; gene therapy;
KW antiatherosclerotic; antiparasitic; virucide; AIDS; Chediak-Higashi syndrome;
KW autoimmune disorder; inflammatory disease; SCID; severe combined immunodeficiency disease; SCID; Chediak-Higashi syndrome;
KW Cushing's disease; Addison's disease; Goodpasture's syndrome; infection;
KW Crohn's disease; diabetes mellitus; Goodpasture's syndrome; infection;
KW Grave's disease; Hashimoto's thyroiditis; Sjogren's syndrome; cancer;
KW Werner's syndrome; cell proliferative disorder; arteriosclerosis;
KW atherosclerosis; cirrhosis; hepatitis; psoriasis.

Human protease and protease inhibitor PPIM-4.

Human; protease; protease inhibitor; protease and protease inhibitor;
PPIM; identification; diagnosis; anti-human immunodeficiency virus; HIV;
antidiabetic; immunostimulant; immunomodulator; antiinflammatory;
antithyroid; immunosuppressive; nephrotropic; antitumor; thyromimetic;
cytostatic; antibacterial; fungicide; protozoicide; hepatotropic; gene therapy;
antiatherosclerotic; antiparasitic; virucide; AIDS; Chediak-Higashi syndrome;
autoimmune disorder; inflammatory disease; SCID; severe combined immunodeficiency disease; SCID; Chediak-Higashi syndrome;
Cushing's disease; Addison's disease; Goodpasture's syndrome; infection;
Crohn's disease; diabetes mellitus; Goodpasture's syndrome; infection;
Grave's disease; Hashimoto's thyroiditis; Sjogren's syndrome; cancer;
Werner's syndrome; cell proliferative disorder; arteriosclerosis;
atherosclerosis; cirrhosis; hepatitis; psoriasis.

Homo sapiens.

WO200110903-A2.

15-FEB-2001.

09-AUG-2000; 2000WO-US021878.

09-AUG-1999; 99US-0147986P.

21-OCT-1999; 99US-0160807P.

(INCY-) INCYTE GENOMICS INC.

Yue H, Lal P, Tang YT, Bandman O, Baughn MR, Azimzai Y, Lu DAM;
Yang J;
WPI; 2001-202760/20.
N-PSDB; AAF81717.

Human protease (inhibitors) useful for diagnosis and treatment of

autoimmune/inflammatory disorders such as acquired immunodeficiency
syndrome, Cushing's disease, Addison's disease and cell proliferative
disorders such as cancer.

Claim 1; Page 95-96; 13app; English.

AA981714 to AAF81740 encode the human proteases and protease inhibitors (PPIMs) given in AAB74668 to AAB74694. The PPIMs can have activities such as: anti-human immunodeficiency virus (HIV); antidiabetic; antichytoid; immunostimulant; immunomodulator; antiinflammatory; immunosuppressive; nephrotropic; antigout; thyromimetic; cytostatic; antibacterial; fungicide; protozoacide; antiatherosclerotic; antipolynucleotide and virucide; antipsoriatic; and hepatotropic. PPIM polynucleotide and protein sequences can be used in the diagnosis, treatment and prevention of autoimmune/inflammatory disorders such as AIDS, DiGeorge's syndrome, severe combined immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's disease, Addison's disease, autoimmune thyroiditis, Crohn's disease, diabetes mellitus, Good pasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and cell proliferative disorder such as arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis and cancer. PPIM polynucleotide sequences can be used in somatic or germline gene therapy and in diagnosis of diseases. They can also be used in generating hybridisation probes useful in mapping the naturally occurring genomic sequences and in molecular biology techniques

XX Sequence 775 AA;

Query Match 100.0%; Score 4070; DB 4; Length 775;
Best Local Similarity 100.0%; Pred. No. 1.8e-310;
Matches 775; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MTIVDKASESDPSAYQNGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
1 MTIVDKASESDPSAYQNGSSEAVSPGDMAGSAGVSSINDVSNHTLSLGPVPGA 60
61 VVYSSSVDPKSPKPOQDQALGDIAPPQKVLFPSEKI CLKWQTHRVGAGLQNLGNTC 120
61 VVYSSSVDPKSPKPOQDQALGDIAPPQKVLFPSEKI CLKWQTHRVGAGLQNLGNTC 120
121 FANAALQCLTYTPPLANYMLSHESKTCBAEGFCMCTMQAHITQALSNPGDVIKPMFVI 180
121 FANAALQCLTYTPPLANYMLSHESKTCBAEGFCMCTMQAHITQALSNPGDVIKPMFVI 180
181 NEMRRIARHLRFGNOEADHEFLQTVDMQKACLSNGSKLDRHTQATTLVQIFGGYLRS 240
181 NEMRRIARHLRFGNOEADHEFLQTVDMQKACLSNGSKLDRHTQATTLVQIFGGYLRS 240
241 RVKCLNCKGVSDTFDYLDTLETKAQS VNKALEQFVKPEQLDGENSYKCSKCKMYPVA 300
241 RVKCLNCKGVSDTFDYLDTLETKAQS VNKALEQFVKPEQLDGENSYKCSKCKMYPVA 300
301 SKRFTIHRSSNVLTLSLKRANFTGKIAKDVKYPEYLDIRPMSQPNGBPIVIVLYAVL 360
301 SKRFTIHRSSNVLTLSLKRANFTGKIAKDVKYPEYLDIRPMSQPNGBPIVIVLYAVL 360
361 VHTGNCAGHYFCVYKASGLWYQNDSTVSTDSIRSVLSQAAVLFYIRSHDVNGGE 420
361 VHTGNCAGHYFCVYKASGLWYQNDSTVSTDSIRSVLSQAAVLFYIRSHDVNGGE 420
421 LTHPTHSQSSPRPVISQVVVNTKQAAPGFIGPOLPSHMIKNPPLHNGTGPKLDTSPSS 480
421 LTHPTHSQSSPRPVISQVVVNTKQAAPGFIGPOLPSHMIKNPPLHNGTGPKLDTSPSS 480
481 MSSPNGNSVNRASPNVASVONMSVNRSSVIPHPKKQKITIISHNKLPRVOCOSQPN 540
481 MSSPNGNSVNRASPNVASVONMSVNRSSVIPHPKKQKITIISHNKLPRVOCOSQPN 540
541 LHSNLENTPKVPSTTNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
541 LHSNLENTPKVPSTTNSAVQSTNSASTMSVSSKVTKPIPRSSCSQPVNMGSKLN 600
601 SSVLVYPVGAESSDSDESKGLGKENGIGTIVSSHSPGQDAEDATPHELPQMTLNGA 660
601 SSVLVYPVGAESSDSDESKGLGKENGIGTIVSSHSPGQDAEDATPHELPQMTLNGA 660

661 NSADSDSPKENGAFDCASQCGQPALHSENPFAKANGLPGLMPAPLLSLPDKILETF 720
661 NSADSDSPKENGAFDCASQCGQPALHSENPFAKANGLPGLMPAPLLSLPDKILETF 720

721 RLSNKLKSTDEMSAPGAERGPPEDRDAEPQPSAAESLEPPDAAASLFFPSEF 775
721 RLSNKLKSTDEMSAPGAERGPPEDRDAEPQPSAAESLEPPDAAASLFFPSEF 775

RESULT 2

AAU98889
ID AAU98889 standard; protease; 1251 AA.

XX AAU98889;

DT 26-AUG-2002 (first entry)

XX Human protease PR75.

XX Protease; human; gastrointestinal disorder; gastritis; atherosclerosis;
XX ulcerative colitis; Reye's syndrome; cardiovascular disorder; enzyme;
XX hypertension; myocardial infarction; autoimmune disease; AIDS; PR75;
XX inflammatory disorder; acquired immunodeficiency syndrome; asthma;
XX Grave's disease; cell proliferative disorder; hepatitis; psoriasis;
XX leukaemia; developmental disorder; Cushing's syndrome; impotence;
XX epithelial disorder; dermatitis; scabies; eczema; neurological disorder;
XX Parkinson's disease; dementia; Alzheimer's disease; infertility;
XX Huntington's disease; multiple sclerosis; reproductive disorder.

XX Homo sapiens.

XX WC200238744-A2.

XX 16-MAY-2002.

XX 18-OCT-2001; 2001WO-US051034.

XX 18-OCT-2000; 2000US-0241573P.

XX 25-OCT-2000; 2000US-0243643P.

XX 02-NOV-2000; 2000US-0245256P.

XX 13-NOV-2000; 2000US-0246395P.

XX 16-NOV-2000; 2000US-0249826P.

XX 20-NOV-2000; 2000US-0252303P.

XX 01-DEC-2000; 2000US-0250981P.

XX (INCY-) INCYTE GENOMICS INC.

XX Lee EA, Hafalia AJA, Yue H, Lal PG, Yao MG, Lu Y, Wallia NK;

XX Warren BA, Lu DAM, Baughn MR, Deleage AM, Burford N, Borowsky ML;

XX Lee S, Xu Y, Griffin JA, Kalliock DA, Gandhi AR, Arvizu C, Ison CH;

XX Tang Y, Azimzai Y, Elliott VS, Swarnakar A, Ramkumar J, Nguyen DB;

XX Tribouley CM, Lo TP, Au-Young J, Thangavelu K, Kearney L;

XX WPI; 2002-463471/49.

XX N-PSDB; ABK86140.

XX New human proteases useful for diagnosing, preventing or treating

XX anorexia, myocardial infarction, Addison's disease, Cushing's

XX syndrome, eczema, Parkinson's disease, and impotence.

XX Claim 1; Page 140-143; 168pp; English.

XX This invention relates to the DNA and protein sequences of novel isolated

XX human proteases (PR75). The protein sequences of the invention are useful

XX for screening a compound for effectiveness as an agonist or antagonist of

XX its activity, the identified agonist and antagonist are useful for

XX treating a disease or condition associated with decreased or

XX overexpression of functional PR75 in a patient. The PR75 protein is also

XX useful as an immunogen for preparing polyclonal or monoclonal antibodies

XX by hybridoma technology. An antibody that binds the PR75 proteins is

XX useful for detection and purification of the proteins and can be used to

XX diagnose a condition or disease associated with expression of PR75 in a

XX subject or in a biological sample. The sequences of the invention are

CC useful for diagnosis, treatment and prevention of gastrointestinal
 CC disorders such as gastritis, ulcerative colitis, Reye's syndrome, etc;
 CC cardiovascular such as atherosclerosis, hypertension, myocardial
 CC infarction, etc; autoimmune/inflammatory disorders such as acquired
 CC immunodeficiency syndrome (AIDS), asthma, Grave's disease, etc; cell
 CC proliferative disorders such as hepatitis, psoriasis, leukaemia etc;
 CC developmental disorders such as Cushing's syndrome; epithelial disorders
 CC such as dermatitis, scabies, eczema, etc; neurological disorders such as
 CC Parkinson's disease, dementia, Alzheimer's disease, Huntington's disease,
 CC multiple sclerosis, etc; or reproductive disorders such as infertility,
 CC impotence. The present sequence represents the PRTS7 protein sequence of
 CC the invention
 XX

SQ Sequence 1251 AA;

Query Match 99.0%; Score 4028; DB 5; Length 1251;
 Best Local Similarity 99.7%; Pred. No. 7.4e-307;
 Matches 767; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MTIVDKASESSDPSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 1 MTIVDKASESSDPSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
 DB |||||
 QY 61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
 DB |||||
 QY 121 FANAALQCLTYTPPLANVMLSHESKTHCAEGFCMCTQOAHITQALSNGPDVVKPMFVI 180
 DB |||||
 QY 121 FANAALQCLTYTPPLANVMLSHESKTHCAEGFCMCTQOAHITQALSNGPDVVKPMFVI 180
 DB |||||
 QY 181 NEMRRIARHLRFNGBDAHEFLQYTVDMOKACLNGSNKLDHRTQATTLVCOIFGGYLS 240
 DB |||||
 QY 181 NEMRRIARHFRGNQEDAEHFLQYTVDMOKACLNGSNKLDHRTQATTLVCOIFGGYLS 240
 DB |||||
 QY 241 RVKCLNCKGVSTDFPYDLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKKWPA 300
 DB |||||
 QY 241 RVKCLNCKGVSTDFPYDLITLLEIKAAQSVNKALEQFVKPEQLDGENSYKSKCKKWPA 300
 DB |||||
 QY 301 SKRFTIHRSSNVLTLSLKEFANFTGKIAKVYPEYLDIRPMSQNPGEIVVLYAVL 360
 DB |||||
 QY 301 SKRFTIHRSSNVLTLSLKEFANFTGKIAKVYPEYLDIRPMSQNPGEIVVLYAVL 360
 DB |||||
 QY 361 VHTGFNCHAGHYFCYIKASNLWYQWDSIVSTSDIRSVLSQAVVLFVIRSHDVKNGE 420
 DB |||||
 QY 361 VHTGFNCHAGHYFCYIKASNLWYQWDSIVSTSDIRSVLSQAVVLFVIRSHDVKNGE 420
 DB |||||
 QY 421 LTHPTHSPQSSPRVVISQVVVNKQAAAPGFIGPQLPSHMIKNPPLHNGTGPKDTPSSS 480
 DB |||||
 QY 421 LTHPTHSPQSSPRVVISQVVVNKQAAAPGFIGPQLPSHMIKNPPLHNGTGPKDTPSSS 480
 DB |||||
 QY 481 MSSPNGNSVNRASPVNASVQWNSVNRSSVVIPEHPKKQKTIISHNKLPVRQCSQPN 540
 DB |||||
 QY 481 MSSPNGNSVNRASPVNASVQWNSVNRSSVVIPEHPKKQKTIISHNKLPVRQCSQPN 540
 DB |||||
 QY 541 LHSNSLENPTKVPVPSSTITNSAVQSTSNASTMSVSKVTKPIPRSESCSQPMNCKSKLN 600
 DB |||||
 QY 541 LHSNSLENPTKVPVPSSTITNSAVQSTSNASTMSVSKVTKPIPRSESCSQPMNCKSKLN 600
 DB |||||
 QY 601 SSVLPVYGAESSEDDESKGLGKENGIGTIVSSHSPQDADEEATPHELQEPMTLNGA 660
 DB |||||
 QY 601 SSVLPVYGAESSEDDESKGLGKENGIGTIVSSHSPQDADEEATPHELQEPMTLNGA 660
 DB |||||
 QY 661 NSADSDPKENGLAPDGCQGPALHSENPFAKANGLPGLKMPAPLLSLPEDKILTF 720
 DB |||||
 QY 661 NSADSDPKENGLAPDGCQGPALHSENPFAKANGLPGLKMPAPLLSLPEDKILTF 720
 DB |||||
 QY 721 RLSNKLKGTDEMAGPAGRGPEDRDPAEPQPGSPAESLEEDPAAASL 769
 DB |||||
 QY 721 RLSNKLKGTDEMAGPAGRGPEDRDPAEPQPGSPAESLEEDPAAAGL 769
 DB |||||

RESULT 3

AAB95146
 ID AAB95146 standard; protein; 1197 AA.
 XX
 AC AAB95146;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human protein sequence SEQ ID NO:17169.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN EF1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 WIPI; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.
 PT
 PT
 XX
 PS Claim 8; SEQ ID NO 17169; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention
 XX
 SQ Sequence 1197 AA;

Query Match 98.8%; Score 4022; DB 4; Length 1197;
 Best Local Similarity 99.6%; Pred. No. 2e-306;
 Matches 766; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MTIVDKASESSDPSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 1 MTIVDKASESSDPSAYQNPQGSSEAVSPGDMAGSAGVSSLNDVSNHTLSLGPVPGA 60
 DB |||||
 QY 61 VYSSSSVDPKSPQKQDQALGDGIAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTC 120
 DB |||||

Db	61	VYSSSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLKMQQTHRVAGLQNLGNTC	120
Qy	121	FANAALQCLTYTPPLANYMLSHSHKTCCHAEFCMCTMQAHITQALSNGPDVVKPMFVI	180
Db	121	FANAALQCLTYTPPLANYMLSHSHKTCCHAEFCMCTMQAHITQALSNGPDVVKPMFVI	180
Qy	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCQIFGGYLRS	240
Db	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCQIFGGYLRS	240
Qy	241	RVKCLNCKGVSDTFPYLDITLTKAOSVKNKALEQFVKPEQDQDGENSYKSKCKKWVPA	300
Db	241	RVKCLNCKGVSDTFPYLDITLTKAOSVKNKALEQFVKPEQDQDGENSYKSKCKKWVPA	300
Qy	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Db	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Qy	361	VHTGFNCCHAGHYFCYIKASNGLWYQMNDSIVSTDSIRSVLSQAYVLYFIRSHDVKNNGE	420
Db	361	VHTGFNCCHAGHYFCYIKASNGLWYQMNDSIVSTDSIRSVLSQAYVLYFIRSHDVKNNGE	420
Qy	421	LTHPTSPGSSRPVTSQRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS	480
Db	421	LTHPTSPGSSRPVTSQRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS	480
Qy	481	MSSPNGNSSVNRASPVNASQVNSVNRSSVPIPEHPKKQKITISIHNLKLPVRCQSQPN	540
Db	481	MSSPNGNSSVNRASPVNASQVNSVNRSSVPIPEHPKKQKITISIHNLKLPVRCQSQPN	540
Qy	541	LHNSLENPKVPSSITNSAVOSTNSASTMSVSKVTKPIPRESESCOPVWNGSKLN	600
Db	541	LHNSLENPKVPSSITNSAVOSTNSASTMSVSKVTKPIPRESESCOPVWNGSKLN	600
Qy	601	SSVLVPYGASSEDSDSESKGLKENGIGTIVSHSPQDAEDDEATPHEIQEPTLNGA	660
Db	601	SSVLVPYGASSEDSDSESKGLKENGIGTIVSHSPQDAEDDEATPHEIQEPTLNGA	660
Qy	661	NSADSDDPKENGAPDASCOQOPALHSENPFAKANGLPKMLPAPLLSLPDKILETF	720
Db	661	NSADSDDPKENGAPDASCOQOPALHSENPFAKANGLPKMLPAPLLSLPDKILETF	720
Qy	721	RLSNKLGSTDEMSAPGAEPPEDRDAPQGPSPAAESLEEDPAAASL	769
Db	721	RLSNKLGSTDEMSAPGAEPPEDRDAPQGPSPAAESLEEDPAAASL	769
RESULT 4			
ID	ABP69291	standard; protein; 1270 AA.	
XX	AC	ABP69291;	
XX	AC	ABP69291;	
DT	20-JAN-2003	(first entry)	
DE	DE	Human polypeptide SEQ ID NO 1338.	
XX	KW	Human; genome mapping; gene therapy; food supplement; virus; fungus;	
XX	KW	cell-proliferative disorder; neurodegenerative disease; bacterial;	
XX	KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;	
XX	KW	multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;	
XX	KW	arthritis; cytosolic; immunomodulator; neurotropic; neuroprotective;	
XX	KW	antiparkinsonian; antidiabetic; immunosuppressive; dermatological;	
XX	KW	haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;	
XX	OS	antiarthritic.	
XX	OS	Homo sapiens.	
XX	PN	WO200270539-A2.	
XX	PD	12-SEP-2002.	
XX	PD		

PF	05-MAR-2002;	2002W0-US003095.	
XX	05-MAR-2001;	2001US-00799451.	
XX	(HYSE-) HYSEQ INC.		
XX	Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;		
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;		
PI	Wehrman T, Wang J, Wang D, Drmanac RT;		
XX	WPI; 2002-759812/82.		
DR	N-PSDB; AB211508.		
XX	New polynucleotides comprising sequences assembled from expressed		
PT	sequence tags (ESTs), useful for treating cell-proliferative,		
PT	neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet		
PT	or coagulation disorders.		
XX	Claim 9; SEQ ID NO 1338; 1012pp + Sequence Listing; English.		
XX	The invention relates to an isolated polynucleotide (I) comprising a		
CC	nucleotide sequence selected from any of 948 sequences (AB211119-		
CC	AB212066) or their mature protein coding portion, active domain coding		
CC	protein or complementary sequences. The polynucleotides are useful for		
CC	identifying expressed genes or for physical mapping of human genome. The		
CC	encoded polypeptides (AEP6902-ABP69849) are useful as molecular weight		
CC	markers, as a food supplement, for generating antibodies, in medical		
CC	imaging, screening and diagnostic assays and for treating cell-		
CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's		
CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,		
CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,		
CC	platelet or coagulation disorders, wound, burns, incision, ulcers, liver		
CC	or lung fibrosis, infections (bacterial, viral, fungal, parasitic),		
CC	arthritis, etc. Note: The sequence data for this patent did not form part		
CC	of the printed specification, but was obtained in electronic format		
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences		
XX	Sequence 1270 AA;		
Qy	Query Match	98.8%; Score 4021; DB 5; Length 1270;	
Db	Best Local Similarity	99.6%; Pred. No. 2.7e-306;	
Qy	Matches	766; Conservative 0; Mismatches 3; Indels 0; Gaps 0;	
Qy	1	MTIYDKASESDPSAYQNPQGSSEAVSPGDMAGSAGVSSLDVSNHTLSLGPVGA	60
Db	1	MTIYDKASESDPSAYQNPQGSSEAVSPGDMAGSAGVSSLDVSNHTLSLGPVGA	60
Qy	61	VYSSSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLKMQQTHRVAGLQNLGNTC	120
Db	61	VYSSSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLKMQQTHRVAGLQNLGNTC	120
Qy	121	FANAALQCLTYTPPLANYMLSHSHKTCCHAEFCMCTMQAHITQALSNGPDVVKPMFVI	180
Db	121	FANAALQCLTYTPPLANYMLSHSHKTCCHAEFCMCTMQAHITQALSNGPDVVKPMFVI	180
Qy	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCQIFGGYLRS	240
Db	181	NEMRRIARHLRFQNGQDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCQIFGGYLRS	240
Qy	241	RVKCLNCKGVSDTFPYLDITLTKAOSVKNKALEQFVKPEQDQDGENSYKSKCKKWVPA	300
Db	241	RVKCLNCKGVSDTFPYLDITLTKAOSVKNKALEQFVKPEQDQDGENSYKSKCKKWVPA	300
Qy	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Db	301	SKRFTIHRSSNVLTLSLKRANFTGGKIAKDVKYPEYLDIRPYMSQPNGEPIVVLYAVL	360
Qy	361	VHTGFNCCHAGHYFCYIKASNGLWYQMNDSIVSTDSIRSVLSQAYVLYFIRSHDVKNNGE	420
Db	361	VHTGFNCCHAGHYFCYIKASNGLWYQMNDSIVSTDSIRSVLSQAYVLYFIRSHDVKNNGE	420
Qy	421	LTHPTSPGSSRPVTSQRVWTKQAAPGFIGPQLPSHMIKNPPLHNGTGPKLDTSPSS	480

Db 421 LTHPTSPGQSSRPVTSQRVTNKQAAPGIFGQLPSHMKNPPLHNGTGLKDTTPSS 480
QY 481 MSSPNGSSVNRASPVNASVQNVSNRSVTPHPPKKQKITIISHNKLPVRQCQSQPN 540
Db 481 MSSPNGSSVNRASPVNASVQNVSNRSVTPHPPKKQKITIISHNKLPVRQCQSQPN 540
QY 541 LHSNSLENPTKVPSSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSQPVNMGSKLN 600
Db 541 LHSNSLENPTKVPSSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSQPVNMGSKLN 600
QY 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITVSSHSPGQDAEDEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITVSSHSPGQDAEDEATPHELQEPMTLNGA 660
QY 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
Db 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKGSTDEMSAPGAERGPPEDRDAEPQGPSAAESLEEPPDAAASL 769
Db 721 RLSNKLKGSTDEMSAPGAERGPPEDRDAEPQGPSAAESLEEPPDAAASL 769
RESULT 5
AAB82177
ID AAB82177 standard; protein; 762 AA.
AC AAB82177;
XX
DT 20-JUL-2001 (first entry)
DE Human ubiquitin protease 23431.
KW Human; ubiquitin protease; deubiquitinating enzyme; tuberculostatic;
KW antiasthmatic; antiinflammatory; antidiarrhoeic; hepatotropic;
KW gynaecological; cytostatic; antimicrobial; neuroprotective; anti-HIV;
KW immunosuppressive; cardiant; antianaemic; nephrotropic; antibacterial;
KW anti-thyroid; gastrointestinal.
XX
OS Homo sapiens.
XX
EN WO200123589-A2.
XX
PD 05-APR-2001.
XX
PF 29-SEP-2000; 2000WO-US026962.
XX
PR 29-SEP-1999; 99US-00407356.
XX
XX (MILL-) MILLENNIUM PHARM INC.
FA Kapeller-Libermann R;
XX
PI
XX
DR WPI; 2001-374253/39.
DR N-PSDB; AAH19322.
XX
XX
PT New human ubiquitin protease, a member of the mammalian deubiquitinating
PT enzymes is useful for diagnosis and treatment of e.g. tuberculosis and
PT Alzheimer's disease.
XX
XX Claim 5; Fig 1; 111pp; English.
XX
XX The present sequence is human ubiquitin protease 23431. The ubiquitin
XX protease coding sequence and protein are useful as reagents or targets in
XX assays for treatment and diagnosis of ubiquitin-mediated or -related
XX disorders, especially disorders mediated by deubiquitinating enzymes. The
XX protein and coding sequence are also useful for treating disorders
XX involving the following: the spleen e.g. tuberculosis and typhoid fever,
XX the lung such as bronchial asthma, the colon e.g. as diarrhoea and
XX dysentery, the liver e.g. jaundice and cholestasis, the uterus and
XX endometrium e.g. endometriosis, the brain e.g. acute meningitis and
XX Alzheimer disease, T-cells including transplant rejection and autoimmune
XX diseases such as systemic lupus erythematosus, diseases of the skin such

CC as malignant melanoma, the bone marrow e.g. B- and T-lymphoid leukaemias,
CC the heart including myocardial infarction, red cells e.g. anaemia, the
CC thymus e.g. Hodgkin disease, B-cells e.g. peripheral B-cell neoplasms,
CC the kidney e.g. polycystic kidney disease, the breast including
CC periductal mastitis, the testis and epididymis e.g. syphilis, the
CC prostate e.g. nodular hyperplasia, the thyroid, e.g. hyperthyroidism, the
CC skeletal muscle e.g. rhabdomyosarcoma, the pancreas e.g. ectopic
CC pancreas, reduced platelet number e.g. HIV (human immunodeficiency virus)
CC -associated thrombocytopaenia and disorders involving precursor T-cell
CC neoplasms including precursor T lymphoblastic leukaemia/lymphoma
XX
SQ Sequence 762 AA;

Query Match 98.3%; Score 3999; DB 4; Length 762;
Best Local Similarity 99.9%; Pred. No. 6.6e-305;
Matches 761; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSASWGAVALNSNHTLSLGPVPGA 60
Db 1 MTIVDKASSSDPSAYONQPGSEAVSPGDMAGSASWGAVALNSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPDKSPKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNGTC 120
Db 61 VVYSSSVDPDKSPKQDQALGDGIAPPQKVLFPSEKICLKWQOQTHRVGAGLQNGTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHESKTCHEAGFCWMCCTMQAHTQALSNGPDVVKPMFVI 180
Db 121 FANAALQCLTYTTPPLANYMLSHESKTCHEAGFCWMCCTMQAHTQALSNGPDVVKPMFVI 180
QY 181 NEMRRIARHLREFGQEDAEHFLQYTVDMQKACLNGSNKLDHRTQATTILVCQIFGGYLRS 240
Db 181 NEMRRIARHLREFGQEDAEHFLQYTVDMQKACLNGSNKLDHRTQATTILVCQIFGGYLRS 240
QY 241 RVKCLNCKGVSDFDPYDITLLEIKAAQSVNKALQFVKPEQLDGENSKYCKCKKWPA 300
Db 241 RVKCLNCKGVSDFDPYDITLLEIKAAQSVNKALQFVKPEQLDGENSKYCKCKKWPA 300
QY 301 SKRFTIHRSSNVLTLSLKRFAFTGCKIAKDVKPYEYLDIRPYMSQNPGEPIVYVYAVL 360
Db 301 SKRFTIHRSSNVLTLSLKRFAFTGCKIAKDVKPYEYLDIRPYMSQNPGEPIVYVYAVL 360
QY 361 VHTGENCHAGHYFCYIKASNGLWYQNDISIVSTDIRSVLSQOAYVLYRSHDVKNGE 420
Db 361 VHTGENCHAGHYFCYIKASNGLWYQNDISIVSTDIRSVLSQOAYVLYRSHDVKNGE 420
QY 421 LTHPTSPGQSSRPVTSQRVTNKQAAPGIFGQLPSHMKNPPLHNGTGLKDTTPSS 480
Db 421 LTHPTSPGQSSRPVTSQRVTNKQAAPGIFGQLPSHMKNPPLHNGTGLKDTTPSS 480
QY 481 MSSPNGSSVNRASPVNASVQNVSNRSVTPHPPKKQKITIISHNKLPVRQCQSQPN 540
Db 481 MSSPNGSSVNRASPVNASVQNVSNRSVTPHPPKKQKITIISHNKLPVRQCQSQPN 540
QY 541 LHSNSLENPTKVPSSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSQPVNMGSKLN 600
Db 541 LHSNSLENPTKVPSSSTITNSAVQSTNSASTMSVSSVTKPIPRSESCSQPVNMGSKLN 600
QY 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITVSSHSPGQDAEDEATPHELQEPMTLNGA 660
Db 601 SSVLPVYGAESSEDSDEESKGLGKENGIGITVSSHSPGQDAEDEATPHELQEPMTLNGA 660
QY 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
Db 661 NSADSDSDPKENGLAPDGASCOQOPALHSENPFKANGLPGLKMPAPLLSLPDKILETF 720
QY 721 RLSNKLKGSTDEMSAPGAERGPPEDRDAEPQGPSAAESLEE 762
Db 721 RLSNKLKGSTDEMSAPGAERGPPEDRDAEPQGPSAAESLEE 762

RESULT 6
ABU08951
ID ABU08951 standard; protein; 762 AA.

XX AC ABU08951;
XX DT 10-JUN-2003 (first entry)
XX DE Human ubiquitin protease.
XX KW Human; enzyme; ubiquitin protease; spleen disorder; Hodgkin's disease;
KW lung disorder; adult respiratory distress syndrome; colon disorder;
KW inflammatory bowel disease; liver disorder; jaundice; uterine disorder;
KW endometriosis; brain disorder; Alzheimer's disease; T-cell disorder;
KW acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;
KW heart disorder; ischaemic heart disease; blood vessel disorder;
KW atherosclerosis; red blood cell disorder; anaemia; thymus disorder;
KW DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder;
KW polycystic kidney disease; glomerulonephritis; breast disorder; mastitis;
KW testicular disorder; sexually transmitted disease; thyroid disorder;
KW hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;
KW whipple disease; tumour; cancer.
XX OS Homo sapiens.
XX PN US2003037350-A1.
XX PD 20-FEB-2003.
XX PF 05-JUN-2002; 2002US-00163547.
XX PR 29-SEP-1999; 99US-00407356.
XX PR 05-NOV-1999; 99US-00435311.
XX PR 01-FEB-2000; 2000US-00496005.
XX PR 11-FEB-2000; 2000US-0182009P.
XX PR 14-FEB-2000; 2000US-0182408P.
XX PR 28-FEB-2000; 2000US-0185503P.
XX PR 02-NOV-2000; 2000US-00704918.
XX PR 12-FEB-2001; 2001US-00781598.
XX PR 14-FEB-2001; 2001US-00782952.
XX PR 28-FEB-2001; 2001US-00796100.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PE Glucksmann MA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;
XX WPI; 2003-342047/32.
XX N-PSDB; ABX93880.
XX PT New nucleic acid or polypeptide, useful for preparing a composition for
XX diagnosing or treating diseases e.g., tumor.
XX PS Claim 9; Fig 1; 281pp; English.
XX CC The invention relates to a new isolated nucleic acid molecule encoding
XX one of 6 polypeptides (ubiquitin protease, lipase, Dynamain, short chain
XX dehydrogenase, ADAM-TS (a disintegrin and metalloprotease domain protein
XX with thrombospondin (TS) domains) and gamma butyrobetaine-hydroxylase
XX (gamma BBH). Also included are the polypeptide, host cells containing the
XX nucleic acid, an antibody that selectively binds to the polypeptide, a
XX method for producing the polypeptide, a method for detecting the presence
XX of the polypeptide or the nucleic acid in a sample, a method for
XX identifying a compound that binds to the polypeptide, a method for
XX modulating the activity of the polypeptide and a method for identifying a
XX compound that modulates the activity of the polypeptide. The nucleic acid
XX or polypeptide is useful for preparing a composition for diagnosing or
XX treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's
XX disease), lung disorders (e.g. adult respiratory distress syndrome,
XX pulmonary oedema, chronic bronchitis and emphysema), colon disorders
XX (e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),
XX liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver
XX disease), uterine and endometrial disorders (e.g. endometriosis and
XX menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's
XX disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell
XX disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders
XX (e.g. urticaria, dermatitis and lupus erythematosus), heart disorders

CC (e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),
CC blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and
CC Raynaud disease), red blood cell disorders (e.g. anaemia), thymus
CC disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),
CC kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),
CC breast disorders (e.g. mastitis), testicular disorders (e.g. sexually
CC transmitted diseases and cryptorchidism), thyroid disorders (e.g.
CC hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal
CC disorders (e.g. whipple disease), as well as tumours and cancers of the
CC above listed organs/cells. Many more diseases and disorders are listed in
CC the specification. The present sequence represents human ubiquitin
XX protease
XX Sequence 762 AA;
SQ
Query Match 98.3%; Score 3999; DB 6; Length 762;
Best Local Similarity 99.9%; Pred. No. 6.6e-305; Indels 0; Gaps 0;
Matches 761; Conservative 0; Mismatches 1;
QY 1 MTIVDKASESSDPSAYQNOPGSGSEAVSPGMDAGSASGAVSLNDVSNHTLSLGPVPGA 60
DB 1 MTIVDKASESSDPSAYQNOPGSGSEAVSPGMDAGSASGAVSLNDVSNHTLSLGPVPGA 60
QY 61 VVYSSSVDPKSPQKQALGDGIAPPQKVLFPSEKICLKWOOTHRVCGALQNLGNTC 120
DB 61 VVYSSSVDPKSPQKQALGDGIAPPQKVLFPSEKICLKWOOTHRVCGALQNLGNTC 120
QY 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFQWCTMOAHITQALNSPDGVIKPMFVI 180
DB 121 FANAALQCLTYTTPPLANYMLSHEHSKTCBAEGFQWCTMOAHITQALNSPDGVIKPMFVI 180
QY 181 NEMRRIARHLRFQNGEDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCOIFGGVYLS 240
DB 181 NEMRRIARHLRFQNGEDAHEFLQYTVDMQKACLNGSKLDRHTQATTLVCOIFGGVYLS 240
QY 241 RVKCLNCKGVSDTFTPYLDITLLEIKAAQSVNKALEQFVKPEQDGENSYKSKCKKWVPA 300
DB 241 RVKCLNCKGVSDTFTPYLDITLLEIKAAQSVNKALEQFVKPEQDGENSYKSKCKKWVPA 300
QY 301 SKRFTHRSSNVLTLILKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPIVTVLYAVL 360
DB 301 SKRFTHRSSNVLTLILKRFANFTGGKIADVKYPEYLDIRPYMSQNGEPIVTVLYAVL 360
QY 361 VHTGFNCHAGHYFCYIKASNGLWYQNMDSIVSTSDIRSVLSQQAYVLYFIRSHDVKNGE 420
DB 361 VHTGFNCHAGHYFCYIKASNGLWYQNMDSIVSTSDIRSVLSQQAYVLYFIRSHDVKNGE 420
QY 421 LTHPTSPGQSSPRPVIISQVVTNKOAPGFIGPOLPSHMIKNPPLHNGTGLKDTSPSS 480
DB 421 LTHPTSPGQSSPRPVIISQVVTNKOAPGFIGPOLPSHMIKNPPLHNGTGLKDTSPSS 480
QY 481 MSSPENGSSVNRASPVNASVQWNSVNRSSVPIPEHPKKQKITIISHNKLPVRCQSQPN 540
DB 481 MSSPENGSSVNRASPVNASVQWNSVNRSSVPIPEHPKKQKITIISHNKLPVRCQSQPN 540
QY 541 LHSNLSLENPTKVPVSSITITNSAVQSTSNASTMSVSSKVTKPIPRSESCSQPMWNGSKLN 600
DB 541 LHSNLSLENPTKVPVSSITITNSAVQSTSNASTMSVSSKVTKPIPRSESCSQPMWNGSKLN 600
QY 601 SSVLVPGAESSESDDESKGLKXENGIGTIVSSHSPGQDAEDEATPHELQEPMTLNGA 660
DB 601 SSVLVPGAESSESDDESKGLKXENGIGTIVSSHSPGQDAEDEATPHELQEPMTLNGA 660
QY 661 NSADSDDPKENGLAPDQASQGOQALHSENPFKANGLPGLKMPAPLLSLPEDKILETF 720
DB 661 NSADSDDPKENGLAPDQASQGOQALHSENPFKANGLPGLKMPAPLLSLPEDKILETF 720
QY 721 RLSNKLKSTDEMSAPGAERGPEDRDAPPGSPAAESLEE 762
DB 721 RLSNKLKSTDEMSAPGAERGPEDRDAPPGSPAAESLEE 762
RESULT 7

ABB06118
ID ABB06118 standard; protein; 558 AA.

XX

AC

XX

DT

XX

DE

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Human NS protein sequence SEQ ID NO:210.
Human; cytostatic; osteopathic; gynaecological; neuroprotective; antiarthritis; antiarthritic; antipsoriatic; ophthalmological; anti-HIV; vasotropic; antiarteriosclerotic; antiinflammatory; dermatological; anorectic; muscular; antiinfertility; cardiovascular; dermatological; antibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiac; anticonvulsant; antidiabetic; tranquilizer; antidiuretic; antilept; gastroenteric; virucide; antiulcer; antidiabetic; antidiuretic; antilept; contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia; endometriosis; degenerative disease; multiple sclerosis; psoriasis; rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma; inflammation; skin disorder; obesity; muscular dystrophy; AIDS; infertility; cardiovascular disease; coagulation disease; hypertension; ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration; diabetes; anxiety; depression; schizophrenia; viral disease; stroke; gastric ulcer; Alzheimer's disease.

Homo sapiens.

WO200206315-A2.

24-JAN-2002.

17-JUL-2001; 2001WO-IL000653.

18-JUL-2000; 2000IL-00137345.

15-DEC-2000; 2000IL-00140354.

(COMP-) COMPUEN LTD.

Mintz L, Freilich S, Bernstein J;

WPI; 2002-155037/20.

N-PSDB; ABL39772.

One hundred and twenty eight novel nucleic acid sequences, useful for

treating and diagnosing e.g. cancer, asthma and Alzheimer's.

Claim 6; Page 243-245; 290pp; English.

ABL39691 to ABL39818 represent novel human nucleic acid sequences

encoding the proteins given in ABB06037 to ABB06164. The novel sequences

(NS) can have cytostatic, osteopathic, gynaecological, neuroprotective,

antirheumatic, antiarthritic, antipsoriatic, ophthalmological, virucide,

vasotropic, antiarteriosclerotic, antiinflammatory, dermatological,

anorectic, muscular, anti-HIV, antiinfertility, cardiovascular,

anticoagulant, antifibrinolytic, hypotension, antiasthmatic, cardiac,

immunomodulator, anticonvulsant, antidiabetic, tranquilizer, antiulcer,

antidepressant, gastrointestinal, antilept, cerebroprotective,

nootropic and contraceptive activities. The NS can be used in vaccines,

gene therapy and antisense therapy. Nucleic acids, expression vectors and

antibodies from the present invention can be used for treating and

diagnosing e.g. cancer, osteoporosis, endometriosis, degenerative

diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,

cataracts, restenosis, atherosclerosis, inflammation, skin disorders,

glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular

disease, coagulation disease, ischaemia, hypertension, asthma, immune

disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,

depression, schizophrenia, viral disease, gastric ulcers, stroke,

Alzheimer's disease and as a contraceptive

Sequence 558 AA;

Query Match 71.4%; Score 2908; DB 5; Length 558;

Best Local Similarity 99.6%; Pred. No. 2.2e-219;

Matches 556; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

209 MQKACLNGSNKLDRTHTQATTLVCOIFGGYLSRSVKCLNCKGVSDFPDYDITLIEIKAAQ 268

1 MQKACLNGSNKLDRTHTQATTLVCOIFGGYLSRSVKCLNCKGVSDFPDYDITLIEIKAAQ 60

269 SVNKALEQFVKPQLDGENSYKCSKCKOMVPASPKRFTIHRSSNVLTLSLRFANFTGGKI 328

61 SVNKALEQFVKPQLDGENSYKCSKCKOMVPASPKRFTIHRSSNVLTLSLRFANFTGGKI 120

329 AKDVKPEYLDIRPMSQNGEPIVVYLAVLVHTGFNCHAGHYFCVYKASNGLWYQMD 388

121 AKDVKPEYLDIRPMSQNGEPIVVYLAVLVHTGFNCHAGHYFCVYKASNGLWYQMD 180

389 SIIVSTDIRSVLSQQAYVLFYIRSHDVKNKGELTHPTSPGQSSPRPVISORVVTNKQAA 448

181 SIIVSTDIRSVLSQQAYVLFYIRSHDVKNKGELTHPTSPGQSSPRPVISORVVTNKQAA 240

449 PGFIGQLPSHMIKNPPHNLGTGPKLDTFSSMSSPNGNSSVNRASPVNASASVQNSVN 508

241 PGFIGQLPSHMIKNPPHNLGTGPKLDTFSSMSSPNGNSSVNRASPVNASASVQNSVN 300

509 RSSVPIPEHKKOKITISIHNLKLPVRCQSQPNLHNSLENPTKVPSSITITSVOSTSN 568

301 RSSVPIPEHKKOKITISIHNLKLPVRCQSQPNLHNSLENPTKVPSSITITSVOSTSN 360

569 ASTMSVSSKVTKPIPRSESCQPVNMGSKLNSVLVPGAESSEDSDESKGLGKNGI 628

361 ASTMSVSSKVTKPIPRSESCQPVNMGSKLNSVLVPGAESSEDSDESKGLGKNGI 420

629 GTIVSSHSPGQDAEDEATPHELQEPMTLNGANSADSDPKENGILAPGASCGQPALH 688

421 GTIVSSHSPGQDAEDEATPHELQEPMTLNGANSADSDPKENGILAPGASCGQPALH 480

689 SENPEKANGLPGLMPAPLISLPEDKILETFLSNKLGSTDEMSAPGAERGPEDRDA 748

481 SENPEKANGLPGLMPAPLISLPEDKILETFLSNKLGSTDEMSAPGAERGPEDRDA 540

749 EPQGSPPAESLEEDPAA 766

541 EPQGSPPAESLEEDPAA 558

RESULT 8

ADE31066

ID ADE31066 standard; protein; 1087 AA.

XX

AC ADE31066;

XX

DT 29-JAN-2004 (first entry)

DE

XX

XX

XX

Human diagnostic and therapeutic polypeptide (DITHP), SEQ ID NO 198.
diagnostic and therapeutic polynucleotide; dithp; antiarteriosclerotic;
antiinflammatory; cerebroprotective; antilipaeamic; antidiabetic;
immunosuppressive; neuroprotective; nootropic; neuroleptic; tranquilizer;
osteopathic; antiarthritic; antirheumatic; cytostatic; hepatotropic;
virucide; haemostatic; anti-HIV; antithyroid; thyromimetic;
dermatological; antibacterial; fungicide; antiparasitic; anticonvulsant;
thrombolytic; anticoagulant; anorectic; vasotropic; antiulcer;
gene therapy; protein replacement therapy; human.

Homo sapiens.

WO2003062376-A2.

31-JUL-2003.

13-JAN-2003; 2003WO-US001096.

16-JAN-2002; 2002US-0349384P.

17-JAN-2002; 2002US-0349413P.

17-JAN-2002; 2002US-0349946P.

XX (INCY-) INCYTE GENOMICS INC.

PA Jones AL, Dahl CR, Gietzen D, Chinn J, Dufour GE, Jackson JL;

PI Yu JY, Tuason O, Yap PE, Amshay SR, Dam TC, Liu TF, Gerstin EH;

PI Peralta CH, Lewis SA, Chen A, Marwaha R, Lan RY, Urashka ME;

PI Kristnam SR, Kolluru V, Panesar IS;

XX WPI: 2003-636732/60.

DR N-PSDB; ADE31255.

XX New human diagnostic and therapeutic polynucleotides and polypeptides,

XX useful for diagnosing, treating or preventing e.g. leukemia, brain

PT cancer, atherosclerosis, AIDS, thyroiditis, infections, obesity, stroke

PT or Alzheimer's.

XX Claim 27; SEQ ID NO 198; 634pp; English.

XX The invention relates to a novel isolated human diagnostic and

CC therapeutic polynucleotide (designated dithp). The novel dithp

CC polynucleotide comprises: any of 188 DNA sequences consisting of 195-7798

CC base pairs fully defined in the specification; a polynucleotide

CC comprising a naturally occurring polynucleotide sequence at least 90%

CC identical to the dithp polynucleotide; a polynucleotide complementary to

CC the dithp polynucleotide or its polynucleotide which is at least 90%

CC identical; or an RNA equivalent of any of the polynucleotides mentioned

CC above. The dithp polynucleotides have the following activities:

CC antiarteriosclerotic, antiinflammatory, cerebroprotective, antilipidemic,

CC antidiabetic, immunosuppressive, neuroprotective, nootropic, neuroleptic,

CC tranquilizer, osteopathic, antiarthritic, antirheumatic, cytostatic,

CC hepatotropic, virucide, haemostatic, anti-HIV, antithyroid, thyromimetic,

CC dermatological, antibacterial, fungicide, antiparasitic, anticonvulsant,

CC thrombolytic, anticoagulant, anorectic, vasotropic, and anticancer. The

CC novel DITHP polynucleotides polypeptide can be used in gene therapy and

CC protein replacement therapy. The dithp polynucleotides or DITHP

CC polypeptides are useful for diagnosing, preventing or treating diseases

CC associated with the expression of human molecules. In particular, these

CC diseases include cancers (e.g. adenocarcinoma, leukaemia, melanoma, brain

CC cancer, breast cancer, cervix cancer, bone cancer, liver cancer, lung

CC cancer) or other cell proliferative disorders (e.g. arteriosclerosis,

CC atherosclerosis, cirrhosis, hepatitis, polycythemia vera, primary

CC thrombocytopenia), autoimmune/inflammatory disorders (e.g. AIDS,

CC Addison's disease, thyroiditis, Crohn's disease, Graves' disease,

CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid

CC arthritis, scleroderma, systemic lupus erythematosus), infections (e.g.

CC viral, bacterial, fungal or parasitic infection), developmental disorders

CC (e.g. Cushing's syndrome or epilepsy), endocrine disorders (e.g.

CC thrombosis, hypopituitarism, hypogonadism, gigantism, goiter) metabolic

CC disorders (e.g. hypercholesterolemia, hypoglycaemia, diabetes,

CC hyperlipidaemia, obesity), neurological disorders (e.g. ischaemic

CC cerebrovascular disease, stroke, Alzheimer's disease, Pick's disease,

CC Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease,

CC anxiety, schizophrenia), gastrointestinal disorders (e.g. ulcers),

CC transport disorders (e.g. akinesia or multidrug resistance), or

CC connective tissue disorders (e.g. Paget's disease or rickets). This

CC sequence represents one of the human DITHP polypeptides of the invention.

XX Sequence 1087 AA;

XX Query Match 35.0%; Score 1423.5; DB 7; Length 1087;

XX Best Local Similarity 41.2%; Pred. No. 2.7e-102;

XX Matches 336; Conservative 109; Mismatches 244; Indels 127; Gaps 25;

QY 1 MTIVDKASSESDPSAYQNPQGSSEAVSPGDMAGSAGWAVSSLNDV----- 47

DB 3 MPVIVDKLEAL-----KPGKDSADDGELKLLASSAKVLLQKIEFEPASKSFSYQL 55

QY 48 ----SNHTLSLGPVPGAVVY-SSSSVPDKSPQKDPQALGDIAPPQKVLPESEKICLK 102

DB 56 EALKSKVLLNPKTEGASRHKSGDDPPARRQSEHTYESCGDGPAPQKVLFPFTELSLR 115

QY 103 WOOTHVRVGLNGLNCTCFANALQCLTYTPPLANTYMLSHSKTKCHARGFCNMCTMQAH 162

DB 116 WERFVRVAGLHNLGHTCFLNATIQCLTYTPPLANTYMLSHSKHARSHCHQSFCMLCVMQNH 175

QY 163 ITQALSNECDVTKPMFVINEMERIARHLRFQNOEDAHEFLQYTVDMQKACLNGSKLDR 222

DB 176 IVQAFANGSNAIKPVSFIRDLKIKIARHFRFGQEDAHEFLRYTIDAMQKACLNGCAKLR 235

QY 223 HTQATTLVCQIFGGYLRSRVKCLNCKGVSDTDPDYLDTILETKAAQSVNKALEBQFVKPQ 282

DB 236 QTOATTLVHQIFGGYLRSEVKCSVCKSVSDTDPYDLIALEIRQANIVRALELFVKADV 295

QY 283 LDGENSYKCSKCKMVPASKRTIHRSSNVLTSLKRFANFTGGKIAKDVKPYEYLDIRP 342

DB 296 LSGENAYMCAKCKKYPASKRTIHRSSNVLTSLKRFANFTGGKIKTDVGPFEFLNIIRP 355

QY 343 YMSQNGEPIVVVLYAVLHVTGCHAGHYFVYKASNLGLYQMNDISIVTSDIRSVLSQ 402

DB 356 YMSQNGDPMYGLYAVLVHSGYCHAGHYCYVVKASNGQYQMNDLSLVHSSNVKVLNQ 415

QY 403 QAYVLYIRSHDVYK3GELTHFTHSPG-QSSPRPVISQVRVTNNKQAAPGFIQPO-IPSHM 460

DB 416 QAYVLYLR-----IPGSKSPEGLISR---TGSSSLPG--RESVIPDHS 455

QY 461 IKNPPLHNGTGPV--KOTPSSSSMSSPNGSSVNRASPVNASVQNVKSVNRSSVIEHPK 518

DB 456 KKNIGNGIISPLTGKRGQSGTMKKPHTTEI--GVPISRNGSTLGLKSONGCIPPKLP 513

QY 519 KQKITISIHKLVRQSQPNLHNSLENP-----TKPVPS-----STITNSAVOSTNAST 571

DB 514 GS-----PSPKLSQTPHTMPLIDDPGKKVKKPAPQHPSPRTAQGLPQTSNSN- 562

QY 572 MSVSSKVTKPIRPSSECSQPVMMGKSLNSSLV- PYGAESSEDSDESKLGENGIGT 630

DB 563 ---SSRSGSQRGSDVSLTSPKLLATATANGHLK---GNDESAGLDR-RGSS 614

QY 631 IVSHSPQDAEAEATPHELQBPMLNGANSADSDPKENGLAPDGACQOGPALHSE 690

DB 615 SSPEHSASSDSTKAFQTPRS-----GAAHLCDQSE-----TNC--STAGHSK 654

QY 691 NPPAKANGLPQKLM-----PAPLLSLPDKILETFRLSNKLGSTDEMSAPGAER 740

DB 655 TPFGADSKTVKLKPVLSNNTTTEPATMTGPPPAK-----KLALSAKKASTLWRATG--- 706

QY 741 GPPEDRDAEPQGPSAAESLEEP-----DAAASLFP 771

DB 707 -----NDLRPPPPSPSSD-LTHPMKTSHPVVVASTWP 736

RESULT 9

AAAY71889

ID AAAY71889 standard; protein; 1123 AA.

XX AAAY71889;

AC AAAY71889;

DT 13-JUN-2001 (first entry)

XX Human ubiquitin protease.

XX Human; ubiquitin protease; cytostatic; virucidal; gene therapy; tumour;

KW neoplasia; hepatic injury; cirrhosis; hepatitis; heart failure; rickets;

KW myocardial infarction; skeletal muscle tumour; Grave's disease; atrophy;

KW glomerulonephritis; cryotorchidism; periductal mastitis; Paget's disease;

KW cystic renal dysplasia; polycystic ovarian disease; cretinism; myxoedema;

KW inflammation; obstructive pulmonary disease; osteoporosis; enterocolitis;

KW idiopathic inflammatory bowel disease; bronchial carcinoma; osteomalacia;

KW choriocarcinoma; nodular hyperplasia; metastatic tumour; thyroiditis.

XX Homo sapiens.

OS Homo sapiens.

EH Key Location/Qualifiers

FT Modified-site 13..16

FT /note= "Amidation site"

FT Modified-site 15..18

FT /note= "cAMP and cGMP-dependent protein kinase"

FT	phosphorylation site"	FT	/note= "Protein kinase C phosphorylation site"
FT	18. .21	FT	566. .568
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	31. .33	FT	570. .575
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	75. .78	FT	582. .584
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	83. .90	FT	595. .600
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	85. .90	FT	607. .610
FT	Modified-site	FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	92. .95	FT	609. .614
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	107. .109	FT	613. .616
FT	Modified-site	FT	/note= "Casein kinase II phosphorylation site"
FT	111. .113	FT	623. .625
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	134. .137	FT	629. .631
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	260. .263	FT	644. .647
FT	/note= "Asn is N-glycosylated"	FT	/note= "Asn is N-glycosylated"
FT	312. .314	FT	656. .659
FT	Modified-site	FT	/note= "Casein kinase II phosphorylation site"
FT	313. .316	FT	662. .664
FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"	FT	672. .675
FT	327. .329	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	/note= "Asn is N-glycosylated"
FT	333. .336	FT	673. .676
FT	Modified-site	FT	/note= "Casein kinase II phosphorylation site"
FT	336. .341	FT	692. .694
FT	/note= "N-myristoylation site"	FT	/note= "Protein kinase C phosphorylation site"
FT	338. .345	FT	694. .697
FT	Modified-site	FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	365. .383	FT	703. .706
FT	Region	FT	/note= "Casein kinase II phosphorylation site"
FT	signature"	FT	748. .750
FT	376. .382	FT	/note= "Protein kinase C phosphorylation site"
FT	Region	FT	765. .767
FT	/note= "MHC immunoglobulins and major histocompatibility complex proteins signature"	FT	/note= "Protein kinase C phosphorylation site"
FT	398. .401	FT	807. .810
FT	Modified-site	FT	/note= "Casein kinase II phosphorylation site"
FT	426. .428	FT	809. .811
FT	/note= "Protein kinase C phosphorylation site"	FT	812. .815
FT	453. .455	FT	/note= "cAMP and cGMP-dependent protein kinase phosphorylation site"
FT	Modified-site	FT	841. .844
FT	467. .470	FT	/note= "Amidation site"
FT	Modified-site	FT	865. .867
FT	467. .469	FT	898. .903
FT	/note= "Protein kinase C phosphorylation site"	FT	/note= "Protein kinase C phosphorylation site"
FT	475. .477	FT	911. .913
FT	Modified-site	FT	/note= "N-myristoylation site"
FT	481. .484	FT	952. .954
FT	Modified-site	FT	/note= "Protein kinase C phosphorylation site"
FT	486. .491	FT	965. .967
FT	/note= "Casein kinase II phosphorylation site"	FT	980. .982
FT	492. .495	FT	/note= "Protein kinase C phosphorylation site"
FT	Modified-site	FT	1031. .1038
FT	493. .498	FT	/note= "Tyrosine kinase phosphorylation site"
FT	Modified-site	FT	1034. .1036
FT	515. .517	FT	/note= "Protein kinase C phosphorylation site"
FT	/note= "Protein kinase C phosphorylation site"	FT	1038. .1041
FT	527. .530	FT	/note= "Amidation site"
FT	Modified-site	FT	1067. .1070
FT	532. .535	FT	/note= "Casein kinase II phosphorylation site"
FT	Modified-site	FT	1103. .1105
FT	546. .548	FT	/note= "Protein kinase C phosphorylation site"
FT	/note= "Amidation site"	FT	1120. .1122
FT	552. .557	FT	
FT	Modified-site	FT	
FT	560. .563	FT	
FT	Modified-site	FT	
FT	/note= "Asn is N-glycosylated"	FT	
FT	561. .563	FT	
FT	Modified-site	FT	

FT XX /note= "Protein kinase C phosphorylation site"
 FN WC200123585-A1.
 PD 05-APR-2001.
 XX 29-SEP-2000; 2000WO-US026915.
 XX 30-SEP-1999; 99US-00408865.
 XX (MILL-) MILLENNIUM PHARM INC.
 PA Kapeller-Libermann R;
 PI WPI; 2001-235374/24.
 DR N-PSDB; AAD03459.
 XX Nucleic acids encoding human ubiquitin protease polypeptides, useful for
 PT preventing, diagnosing and treating, e.g. breast, lung or liver cancer
 PT and viral infections.
 PS Claim 9; Fig 1; 120pp; English.
 XX The present sequence is human ubiquitin protease belonging to the family
 CC of mammalian deubiquitinating enzymes. This ubiquitin protease is encoded
 CC by the cDNA insert of plasmid deposited with ATCC NO: PTA-1849. Ubiquitin
 CC protease is highly expressed in foetal kidney, testes, foetal liver, and
 CC ovary and foetal heart. It is also expressed in liver, breast, lung, and
 CC colon tissue and in liver metastasis derived from malignant colon
 CC tissues. Ubiquitin protease is useful for treating disorders mediated by

Query Match 35.0%; Score 1423.5; DB 4; Length 1123;
 Best Local Similarity 37.3%; Pred. No. 2.8e-102;
 Matches 347; Conservative 112; Mismatches 238; Indels 233; Gaps 25;

1 MTIVDKASESDPSAYQNPQSGSEAVSPGDMAGSAGVSSLNDV----- 47
 1 MPVDKLKEAL-----KPRKDSADGELGKLLASGAKVLLQKIEFEPASKSFSYQL 53
 48 ----SNHTLSLGPVPGAVVY-SSSVVPDKSPKPOKDALGDGIAPPOKVLFPSPKICLK 102
 54 EALKSKYVLLNPKTEGASRHKSGDDPPARRQSGEHTYESCGDGPAPQKVLFPPTBELSLR 113
 103 WOOTHVRGAGLQNLGNTCFANALQCLTVTPPLANVMSLSHSKTCCHAEFGCMCTMQAH 162
 114 WERFVRVAGLHNLGNTCFNLATIQCLTVTPPLANVLSKHSRCHQSGFCMLCVMQNH 173
 163 IQQALSNPGDVKIPFVINEMRRIARHLRFGNQEDAHEFLQTVDMQKACLNGSKLDR 222
 174 IVQAFANSNAIKPVSFIRDLKIAHRHFRFGNQEDAHEFLRYTIDAMQKACLNGCAKLR 233
 223 HQQATTLVQIIFGGYLSRVKCLNCKGVSDTDPYLDITLLEIKAAQSVNKALEQFVKPQ 282
 234 QTQATTLVHQIIFGGYLSRVKSVCKSVSDTYDPLDVALEIRQAANIVRALFLFKADV 293
 283 LDGENSYKSCCKMVPASKRFTIHRSSNVLTLSLKRFRANFTGGKIADKVKYPEYLDIRP 342
 294 LSENAVMCAKCKKVPASKRFTIHRSSNVLTLSLKRFRANFSGKIKTDVGYPEFLNIRP 353
 343 YMSQNPGEPIVYVLYAVLVHTGNCCHAGYFCVYKASNLWYQMNDSIVSTSDIRVLSQ 402
 354 YMSQNGDPVMYGLYAVLVHSGYCHAGHYCYCVKASNGQWYQMNDSLVHSSNVKVLNQ 413
 403 QAYVLFIRSHDVKNGE-----LTHPTSPGQSGPRP-----VISQRVYTNKQ-- 446
 414 QAYVLFVLRIPGSKSPGGLISRTGSSSLPGRSVIPDHSKKNIGNGLISPLTGKRQDS 473
 447 -----AAPGFIGQLPS-----HMKVPHLNGTGLPKD 475
 474 GTMKKHTTBEIGVPIRNGSTGLKSGNGCICPKLPSPGSPKLSQTPTHM---PTILD 530
 476 TPSSSMSP-----NGNS-----VNRASPV----- 496

of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 954 AA;

Query Match 35.0%; Score 1422.5; DB 4; Length 954;
Best Local Similarity 41.1%; Pred. No. 2.6e-102; Mismatches 244; Indels 127; Gaps 25;
Matches 335; Conservative 110; Mismatches 244; Indels 127; Gaps 25;

QY 1 MTVDKASSESSPSAYQNGPSSGSEAVSPGMDAGSAGWAVSSINDV----- 47
Db 1 MPVDKLEAL-----KPRKDSADDGELGKLASSAKVLLQKTEPEPASKSPSYQL 53
QY 48 ----SNHTLSLGPVGAVVY-SSSPVDSKSPQKQALGDGIAPPOKVLFPSEKICLK 102
Db 54 EALKSKYVLLNFKTEGASHKSGDDPPARRQSEHTYESCGDGPAPQKVLFPPTERLSLR 113
QY 103 WQTHRVGAGLQNLGNTCPANALQCLTTPPLANTMLSHESKTKCHAGFCWCMCTMQAH 162
Db 114 WERFVRGAGLHNLGNTCFNLATIQLTTPPLANYLLSKHARSCHQGSFCLMCLQMNH 173
QY 163 ITQALSNPDVTKMPFVINEMRIRARHFRFQEDAEHEFLQYTVDMQKACLINGSNKILDR 222
Db 174 IVQAFANGNAIKPVSFIRDLKRIARHFRFQEDAEHEFLRYTIDAMQKACLINGCAKLDR 233
QY 223 HTQATTLVCQIFGGYLSRVKCLNCKGVSDTDPDYLDTLEIKAAQSVNVALEQVFKPEQ 282
Db 234 QTOATLHQIFGGYLSRVKCLNCKGVSDTDPDYLDTLEIKAAQSVNVALEQVFKADV 293
QY 283 LDGENYKSCCKKMPVASKRFTIHRSSNVLTLSLKRFAFTGGKIADVKYVELDTRP 342
Db 294 LSGENAYVCAKCKKVPASKRFTIHRSSNVLTLSLKRFAFTGGKIADVKYVELDTRP 353
QY 343 YMSQNGEPIVYLVYVHTGFCNCHAGYCYTKASNGLWYQMDISVTSDIRSVLSQ 402
Db 354 YMSQNGDPVMYGLYAVLVHSGYCHAGYCYTKASNGWYQMDISLVHSSNVKVLNQ 413
QY 403 QAYVLFYTRSHDVKNGBELTHPTSPG-QSSPRPVISORVTVNQAPGFTGPQ-LPSHM 460
Db 414 QAYVLFYLR-----IPGSKSPEGLISR---TGSSSLPG--RPSVIPDHS 453
QY 461 IKNPPLHNGTGPL--KDTPTSSMSPPNGSNVNRASPVNASQVNWVNRSSVIBHPK 518
Db 454 KKNLNGIISPLTGRQDSGTMKKPHTEI--GVPIRNGSTLGLKQNGCIPPKLPS 511
QY 519 KQKTIISHNKLPRQCSQNLHNSLENP-----TKPVPS---STTNSAVQSTSNAST 571
Db 512 GS-----PSPKLSQTPHTFTILDDPGKKVKPAPQHPSPRTAQLPCTSN 560
QY 572 MSVSSKVTKPIPRESCSQPMNCKSLNSVLV-PYGASSESDSDRESKGLKENGIGT 630
Db 561 ---SSRSGSORQSGWRDVLVTSFKLLATATANGHLK----GNDESAGLDR-RGSSS 612
QY 631 IVSSHSPQDAEDAEATPHELQEPMTLNGANSADSDPKENGLAPDGASCQCPALHSE 690
Db 613 SSPEHSASSDSTKAPQTPRS-----GAHLCDQSE-----TNC--STAGHSK 652
QY 691 NPAKANGLPGKLM-----PAPLLSLPEDKILETFLSNKLGSTDEMSAPGAER 740

Db 653 TPSPGADSKTVKLKSPVLSNNTTTEPASTMTGPPPAK-----KLASAKKASTLWRATG--- 704
QY 741 GPPEDRDAEPQPGSPAASELEP-----DAAASLFP 771
Db 705 -----NDLRPPPPSPSSD-LTHPMKTSHPVVASTWP 734

RESULT 11
AAB95208
ID AAB95208 standard; protein; 548 AA.
XX AAB95208;
AC AAB95208;
DT 26-JUN-2001 (first entry)
XX Human protein sequence SEQ ID NO:17313.
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
KW Homo sapiens.
XX EP1074617-A2.
XX 07-FEB-2001.
XX 28-JUL-2000; 2000EP-00116126.
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

Claim 8; SEQ ID NO 17313; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 548 AA;

Query Match 33.8%; Score 1375; DB 4; Length 548;

Best Local Similarity 47.3%; Pred. No. 6.1e-99; Mismatches 81; Mismatches 152; Indels 86; Gaps 12; Matches 286; Conservative

QY 1 MTIVDKASSESSPPSAVQNPQSGSEAVSPGDMAGSAGVSSSLNDV-----47

DB 1 MPIDVKLKEAL-----KPGKRDGADGELKLLASSAKKVLQKIEPEPASKSPSYQL 53

QY 48 ----SNHTLSLGPVAVVY-SSSSVPDKSPQKQDQALGDIAPPQKVLFPSEKICLK 102

DB 54 EALKKXVLLNPKTEGASRHKSGDDPARRRGSEHTYESCGDGPAPQKVLFPPTERLSLR 113

QY 103 WOQTHRVAGLQNLGNTCFANALQCLTTPPLANVLSHESKTKCHAEFCFMMCTMQAH 162

DB 114 WERVFRVAGLHNLGNTCFNLATIQCLTTPPLANVLSHESKTKCHAEFCFMMCTMQAH 173

QY 163 ITQALSNPGDVLPKMFVINEMERIARHLRFGNEDAEFLQYTDAMQKACLNGSKNLD 222

DB 174 IVQAFANSNAIKPVSFIRDLKKIARHFRFGNEDAEFLRYTIDAMQKACLNGSKNLD 233

QY 223 HTQATLVQCIIFGGYLSRVKCLNCKGVSDFDPYLDITLIEIKAAOSVNKALBQFVKPEQ 282

DB 234 QTAQTLVHQIFGGYLSRVKCSVKSVSDTYDYLDALEIRQAANIVRALEFLFKADV 293

QY 283 LDGENSKSKCKKMPASKRFTIHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRP 342

DB 294 LSGENAYMCAKCKKVPASKRFTIHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRP 353

QY 343 YMSQPNGEPTVYVLYAVLHVTGFGNCHAGHYFCYIKASNLGLWYQWDSIVSTDIRSVLSQ 402

DB 354 YMSQNGDPWYGLYAVLHVSCHAGHYCYVVKASNGQWQWDSIVSTDIRSVLSQ 413

QY 403 QAYVLYFIRSHDVKNGELTHPSPG-QSSPRPVISQVTVTKQAAPGIGPQ-LPSHM 460

DB 414 QAYVLYFIRSHDVKNGELTHPSPG-QSSPRPVISQVTVTKQAAPGIGPQ-LPSHM 453

QY 461 IKNPPLHNGTGL--KDTPLSSKSSPNSGNSVNRASPVNASVQWNSVRSSVPIPEHPK 518

DB 454 KKNIGNGIISPLTGKRDGTMKKPHTTEI--GVFLRNGSTLGLKQNGCIPPKLPS 511

QY 519 KQKTIISHNKLPVRQCSQNLHNSLENFTKVPSPSTITNSAVQSTSNASTMSVSSKV 578

DB 512 GS-----PSPKLSQPTHTMTILDGPK-----KV 536

QY 579 TKPIP 583

DB 537 KKPAP 541

RESULT 12

ADC27023

ID ADC27023 standard; protein; 565 AA.

XX AC ADC27023;

XX DT 18-DEC-2003 (first entry)

XX DE Human deubiquitinating enzyme DUB4.8 SEQ ID NO:20.

XX KW human; deubiquitinating protease; enzyme; hDUB; DUB; autoimmune disease; infection; antiinflammatory; immunosuppressive; hDUB4; chromosome 4.

XX OS Homo sapiens.

XX PN WO2003072724-A2.

XX PD 04-SEP-2003.

XX PF 20-FEB-2003; 2003WO-US005338.

XX PR 22-FEB-2002; 2002US-0358873P.

XX PR 22-FEB-2002; 2002US-0358873P.

XX PR 08-MAR-2002; 2002US-0363020P.

XX PR 12-APR-2002; 2002GB-00008404.

XX (AVET) AVENTIS PHARM INC.

XX Hahn C, Liu H;

XX WPI; 2003-721759/68.

XX N-PSDB; ADC27022.

XX New isolated polynucleotide encoding human deubiquitinating proteases, useful for identifying e.g. antiinflammatory or immunosuppressive agents, also the encoded polypeptides.

XX Claim 2; SEQ ID NO 20; 201pp; English.

XX The present invention describes human deubiquitinating protease enzymes (I), designated hDUB. The enzymes designated hDUB4 are located on chromosome 4, and the enzymes designated hDUB8 are located on chromosome 8. Also described is a method for reducing inflammation, modulating autoimmune diseases and modulating immune reactions during infection by administering a compound (A) that inhibits (I) or alters regulation of the transcription of nucleic acid sequences encoding (I). (I) has antiinflammatory and immunosuppressive activities, and can be used in the regulation of cytokine/cytokine receptor signaling involved in lymphocyte proliferation. (II) can be used for identifying their respective inhibitors (A). (A) are useful for reducing inflammation and to modulate autoimmune diseases or immune reactions during infections. The present sequence represents hDUB4.8 from the present invention.

XX SQ Sequence 565 AA;

Query Match 28.3%; Score 1150.5; DB 7; Length 565;

Best Local Similarity 46.7%; Pred. No. 2.8e-81;

Matches 244; Conservative 78; Mismatches 162; Indels 39; Gaps 8;

QY 28 PGDMAGSAGVAVSLNDVSNHTLSLGPVAVVYSSSSVPDKSPQKQDQALGDI 87

DB 33 PVDMDLSYLGEGWQNFHFSKLT-SRPDAFAEIQTSLPEKSPCETRVLDCCDLA 91

QY 88 PPOKVLPPSEKICLFWOQTHRVAGLQNLGNTCFANALQCLTTPPLANVLSHESK 147

DB 92 PVARQLAPREKLPLSSRPAAVAGLQNLGNTCFANALQCLTTPPLANVLSHESK 151

QY 148 CHAEFCFMMCTMQAHITQALSNPGDVLPKMFVINEMERIARHLRFGNEDAEFLQYTD 207

DB 152 CHRHKGMCLCTMQAHITRALHNPGHVQP-----SQALAGFRHGRQEDAEFLQYTD 205

QY 208 AMQKACLNGSKNLDHPTQATLVQCIIFGGYLSRVKCLNCKGVSDFDPYLDITLIEKAA 267

DB 206 AMKACLPFGHKQVDHHSKDTTLHQIFGGYVRSQIKLCHGIGSDTFDPYLDIALDIAA 265

QY 268 QSVNKALEQFVKPEQLDGENSYKSKCKKMPVASKRFTIHRSSNVLTSLKRFANFTGK 327

DB 266 QSVQALEQLVKPELNGENAYHGVCLQAPASKTLTLHTSAKVLILVLRFSVDTGNK 325

QY 328 TAKDVKYPEYLDIRPYMSQPNGEPTVYVLYAVLHVTGFGNCHAGHYFCYIKASNLGLWYQW 387

DB 326 TAKNVQYPECLDMQPYMSQNTGTPVYVLYAVLHVTGFGNCHAGHYFCYIKASNLGLWYQW 385

QY 388 DSIYSTDIRSVLSQAAYVLFYIR-----SHDVKNKGELTHPHTSPGQSSPRVVISQR 440

DB 386 DAEVTAASITSVLSQAAYVLFYIR-----SHDVKNKGELTHPHTSPGQSSPRVVISQR 443

QY 441 VVTNKQAAPGFI-----GPQLPSHMIKPNPHLNGTGLKDTPTSSSSMSPNGN-SSVN 491

DB 434 --TDRRATQGLKROHPCLQAPDELHILVERATQESTLDHWKFLQEQNKTKPEFNRKVE 491

QY 492 RASPVNASVQWNSVNRSSVPIPEHPKQKTIISHNKLPVRQ 534

DB 492 GTLP--PDVLIHQSKYKGMKNHHPHQSSSLNLSSTPTTHQ 532

RESULT 13

ABU10207

XX (AVET) AVENTIS PHARM INC.
XX Hahn C, Liu H;
XX WPI; 2003-721759/68.
XX N-PSDB; ADC27014.
XX New isolated polynucleotide encoding human deubiquitinating proteases,
XX useful for identifying e.g. antiinflammatory or immunosuppressive agents,
XX also the encoded polypeptides.
XX
XX Claim 2; SEQ ID NO 12; 201pp; English.
XX
XX The present invention describes human deubiquitinating protease enzymes
XX (I), designated hDUB. The enzymes designated hDUB4 are located on
XX chromosome 4, and the enzymes designated hDUB8 are located on chromosome
XX 8. Also described is a method for reducing inflammation, modulating
XX autoimmune diseases and modulating immune reactions during infections by
XX administering a compound (A) that inhibits (I) or alters regulation of
XX the transcription of nucleic acid sequences encoding (I). (I) has
XX antiinflammatory and immunosuppressive activities, and can be used in the
XX regulation of cytokine/cytokine receptor signaling involved in lymphocyte
XX proliferation. (I) can be used for identifying their respective
XX inhibitors (A). (A) are useful for reducing inflammation and to modulate
XX autoimmune diseases or immune reactions during infections. The present
XX sequence represents hDUB4.3 from the present invention.
XX
XX Sequence 530 AA;
XX
XX Query Match 28.0%; Score 1140.5; DB 7; Length 530;
XX Best Local Similarity 46.5%; Pred. No. 1.5e-80;
XX Matches 242; Conservative 78; Mismatches 161; Indels 39; Gaps 8;
XX
Qy 31 MDAGSASGAVSSLDVSNHTLSLGPVGVVYSSSVDPKSPQKQDQALGDGIAPPQ 90
Db 1 MEDDSLVLGGEWQHFHFSKLT-SRPDAAFAEIQTSLPEKSPLCSETRVDLCDDLPVA 59
Qy 91 KVLFPSEKICLKWOQTHRVAGLQNLGNCTCFANAALQCLTTPPLIANTMLSEHSKTCOA 150
Db 60 RLAPREKPLSSRRPAAGVAGLQNMGNCTCVNASLQCLTTPPLIANTMLGREHSQTCR 119
Qy 151 EGFCCMCTMQAHITQALSNPGDVLKMPFVINEMRRIARHFRGQEDAEHFLQYTVDAWQ 210
Db 120 HGGMLCTMQAHITRALNPGHVLP-----SQALAGFRHGRQEDAEHFLMTFTVDAMK 173
Qy 211 KACLNGSNKLDRTQATLVCCQIFGGYLSRVKCLNCKGVSDTFDPPYLDITLEIKAAQSV 270
Db 174 KACLPGHKQVDHSHKDTTLIHIFGGYWRSQIKCLHCHGISDTPDPPYLDIALDICAQSV 233
Qy 271 NKALEQFVKPQLDGENSYKCSKCKMVPASKRFTTHRSSNVLTLSLKRPFANFTGGKIAT 330
Db 234 QQALEQLVKPEELNGENAYHCGVCLQRAPASKTLTLHTSAKVLILVLKRFSDVTGNKIAT 293
Qy 331 DVKYPEYLDIRPYMSQNGEPIVVYLYAVLVHTGPNCHAGHYECYIKASNGLYQMDSI 390
Db 294 NVQYPECLDMQPYMSQNGTGPLVYLYAVLVHAGWSCHNGHYFSYKRAQEGQWYKMDAE 353
Qy 391 VSTDIRSVLSQQAYVLYFIR-----SHDVKNKGELTHTSPGSSPRPVISQRVVT 443
Db 354 VTAASITSVLSQQAYVLYFIQKSEWERHSESVSRGRE-----PRALGAED--T 399
Qy 444 NKQAPGFI-----GPQLPSHMKNPPLHNGTGPLKDTFPSSSMSSPNGN-SSVNRAS 494
Db 400 DRRATQGBELKRDHPCLQAPDELDEHLVERATQESTLDRWKFLQEQNKTKPEFNRKVEGTL 459
Qy 495 PVNASASVQNWSVNRSSVYPEHPKKQKITITSHNKLVPVRO 534
Db 460 P--PDVLVIHOSKYKCGMKNHHPEQQSSLLNLSSSTPTHQ 497

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:56:53 ; Search time 20 Seconds
(without alignments)
3727.422 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 4070
Sequence: 1 MTIVDKASESDPSAYQNP.....AAESLEPDAASLFPFSEG 775

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues 283366
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	846.5	20.8	526	JC6133	deubiquitinating e
2	760	18.7	457	T41069	ubiquitin carboxyl
3	720	17.7	891	H86306	F20D23.20 protein
4	703	17.3	743	B84639	probable ubiquitin
5	692.5	17.0	631	T05103	hypothetical prote
6	626.5	15.4	1008	T05578	hypothetical prote
7	504.5	12.4	1082	T05075	hypothetical prote
8	428	10.5	471	S57591	ubiquitin specific
9	370.5	9.1	557	T50788	ubiquitin carboxyl
10	358.5	8.8	449	T37611	hypothetical prote
11	351.5	8.6	1672	T46237	hypothetical prote
12	343.5	8.4	792	S63141	probable ubiquitin
13	336	8.3	365	B84611	ubiquitin-specific
14	332	8.2	805	S50277	deubiquitinating enz
15	323.5	7.9	371	T06097	hypothetical prote
16	304	7.5	926	S39344	hypothetical prote
17	303.5	7.5	264	T47183	probable ubiquitin
18	296	7.3	408	T40715	probable ubiquitin
19	295.5	7.3	585	T50325	hypothetical prote
20	287.5	7.1	825	T47164	hypothetical prote
21	275	6.8	555	T23531	ubiquitin carboxyl
22	267	6.6	438	T39772	protein T05H10.1
23	265.5	6.5	1292	A98225	hypothetical prote
24	265.5	6.5	1292	T24559	fat facets protein
25	264.5	6.5	2559	T30850	fat facets (faf) s
26	260	6.4	2747	B49132	probable ubiquitin
27	257	6.3	512	T40815	hypothetical prote
28	256	6.3	754	S50601	hypothetical prote
29	254	6.2	1430	T21910	hypothetical prote

30	253	6.2	849	2	T41085	probable ubiquitin
31	251	6.2	1108	2	T41188	probable ubiquitin
32	250.5	6.2	1230	2	S53974	hypothetical prote
33	246.5	6.1	928	2	T04192	hypothetical prote
34	243.5	6.0	937	2	T04194	probable ubiquitin
35	243	6.0	350	2	T39795	probable ubiquitin
36	242.5	6.0	1095	2	T20528	hypothetical prote
37	241	5.9	1130	2	T23104	hypothetical prote
38	241	5.9	1133	2	T23103	hypothetical prote
39	236.5	5.8	887	2	C86453	CDS protein F9111.
40	229.5	5.6	376	2	T30979	hypothetical prote
41	229	5.6	794	2	I58376	hypothetical prote
42	219	5.4	688	2	S45803	ubiquitin thiolest
43	218	5.4	963	2	T09478	ubiquitin-specific
44	217	5.3	912	2	B44450	probable ubiquitin
45	217	5.3	914	2	T00757	probable ubiquitin

ALIGNMENTS

RESULT 1

JC6133
deubiquitinating enzyme - mouse
N:Alternate names: DUB-1 protein
C:Species: Mus musculus (house mouse)
C>Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 05-Nov-1999
C:Accession: JC6133
R:Zhu, Y.; Carroll, M.; Papa, F.R.; Hochstrasser, M.; D'Andrea, A.D.
Proc. Natl. Acad. Sci. U.S.A. 93, 3275-3279, 1996
A:Title: DUB-1, a deubiquitinating enzyme with growth-suppressing activity.
A:Reference number: JC6133; MUID:96194957; PMID:8622927
A:Accession: JC6133
A:Molecule type: mRNA
A:Residues: 1-526 <ZHU>
A:Cross-references: GB:U41636; NID:gl302629; PIDN:AA052532.1; PID:gl302630
C:Comment: This enzyme is the first enzyme of the ubiquitin system directly implicated in and cell cycle progression, and in cytokine-induced cell proliferation.
C:Genetics:
A:Gene: dub-1

Query Match 20.8%; Score 846.5; DB 2; Length 526;
Best Local Similarity 37.1%; Pred. No. 5.9e-43;
Matches 186; Conservative 89; Mismatches 152; Indels 75; Gaps 12;

QY	76	PKDQALGDIAPP-----QKVLFPSEKICLKWOQTHRVGAGLQNLGNTCFANAALQ	127
DB	8	PEADPPALSPDAPELHQDEAQQVVEELTVNGKHSLSWESPGCGGLQNTGNSCYLNAALQ	67
QY	128	CLTYTPPLANYMLSHESHTCHAEFGFCMMCTMOAHITQAL--SNPGDVTKPMFVINEMRR	185
DB	68	CLTHTPPLADYMLSQHSQTCSPGCKLCAMEALVTQSLHSHSGDMVKPSHILTS---	124
QY	186	IARHLRFQNGQDAHEFLQYTVDAQKACLNGSKLDRHTQATTLVQIFGGYLSRVKCL	245
DB	125	-AFHKH--QQEDAHEFLMTLETMHESCLQVHRQSKPTSEDSPIHDFIGGWRSSQIKCL	181
QY	246	NCKGVSDTPPYLDITLLEIKAAQSVNKAQFVKPEQLDCGNSYKSCKCKMVPASKRFT	305
DB	182	LCQGTSDTYDFRIDIPLDISSAQSVKQALWDTEKSELCDGNAYYCGCKQKWPASKTLH	241
QY	306	IHRSSNVLTSLKRFANFTGGKIAKQVYPEYLDIRPYMSQNPGEPIYVYVAVLVHTGF	365
DB	242	VHIAPKVLVNLRFSAFTGNKLDKRVSYPEFLDLKPYLSEPTGGPLPYALYAVLVHDGA	301
QY	366	NCHAGHYFCYIKASNGLWYQNDISIVTSDIRSVLSQAYLVFYIR-----SHDVKNG	418
DB	302	TSHSGHYFCVCKAGHKWYKMDDTKVTRCDVTSVLNENAYLVFYVQANLKQVSDIMPEG	361
QY	419	--GELTHPTH-----SPQSSRPVISO-----RWVTNKQ	446
DB	362	RINEVLDPYQLKRSRRKKKKSPFTEDLGPENCKDKRAIKETSLGKGVLEVNKK	421

A,Reference number: A84420; MUUD:20093487; PMID:10617197
A,Accession: B84639
A,Status: preliminary
A,Molecule type: DNA
A,Residues: 1-743 <STO>
A,Cross-references: GB:AEO02093; NID:94572681; PIDN:AAD2
C,Genetics:
A,Gene: At2g24640
A,Map position: 2

Query Match	17.38;	Score 703;	DB 2;	Length 743;
Best Local Similarity	31.2;	Pred. No. 3.3e-34;		
Matches 189;	Conservative 92;	Mismatches 241;	Indels 84;	Gaps 15;
QY	60	AVVSSSSVPDKSPSPQKQALGDGIAPQKVLPFPSEKIC--LKWQOTHRVRVAGLQNLG	117	
Db	174	ASLPGNESASKTRIALVPQOSQSOKATLKPTDVLFPYESFVRYVYVNDRPTMAPCGLTNCG	233	
QY	118	NTCFANAALQCLTTPPLANYMLSHEHSKTCBAEGFCMCTMQAHITQALNSPGDVIKPM	177	
Db	234	NSCFANVVLQCLSWTRFLVAYLLERGHKRCRRNDWCLCFEFENHLDR--NYSRPFPSPM	292	
QY	178	FVINEMBRIRARHLRFGNOEDAHEFLOVTVDAQKACLN--GSKLDRHTQATTLVCOIF	234	
Db	293	NIISRLPNIGNLGYGEQEDAEHMLRAIDMMQSVCLDFEGGEKVPVPRQETTLIOYIP	352	
QY	235	GGYLRSVKCLNCKGVSDTDPYDITLTKA--AQSVMKALBQFVKPEQLDGENSYKCSK	293	
Db	353	GGLLQSQVQCTACSNVSDQVENMMDLTVEIHGDVAVSLEECLDQFTAKEWLQGDNLVKCDR	412	
QY	294	CKKMWPAKSFTHRSSNVLTLSLKRANFTGGKIAKDVKPYEVLDIRPYMSQFNGEPIV	353	
Db	413	CDYVVKACKRLSIRCAPNILLTALRKFGGRFGKLNKRISFETFDLGPYMSGGEGSDV	472	
QY	354	VYLYAVILVHTGF--NCHAGHYFCYIKASNGLWYQNDPSIVSTSDIRSLSQAYVLYFIR	411	
Db	473	YKLYAVIVHLMDLNASFFGHYICVVKDFRGNWYRIDDSEVKEVELEDVLSQRAYMLLYSR	532	
QY	412	SHDVKNGETL-HPTHSPGGSSPRPVISQRVVTKNQAAAPGIGQLPSHMTKNPPLHNGT	470	
Db	533	T--VLSEGALTDDYDIWYRCVRQPRP-----SNLRSEESQDEKKTDTL	572	
QY	471	GPLKDTPTSSMSSPNGNSVNRASPVNASASQVNWNSRRSVIEPHPKPKQKITTSIHNLK	530	
Db	573	NTESNQGSVESSGVGINDTSTVSSLNGLIIS-----HSEDEYEKESLSASV----	620	
QY	531	PVRQ-----COSQPNL-----HNSLENPTKVPVPSTTTNSAVQ	564	
Db	621	PVSEGEKGVDKVTDVDSSENRSDIMEHDSGTDHQEEFANGKEDP-----TVENLAVD	673	
QY	565	STSNASTMSVSSKVTKPIP-----BSESGQPVNMGKSLNSSLVLPYGAESSESDDESK	620	
Db	674	SSCLDITTPSPSAATEFIPQENERSDTSKPLEKHSDDTESNKPL-----EKEHLDESK	728	
QY	621	GLGKEN	626	
Db	729	PLEXEH	734	

RESULT 5
T05103

T05103
 hypothetical protein F28W20.140 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
 C:Accession: T05103
 R:Byevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Hobeisel, J.; Mewes,
 submitted to the Protein Sequence Database, November 1998
 A:Reference number: Z15398
 A:Accession: T05103
 A:Molecule type: DNA
 A:Residues: 1-631 <BEV>
 A:Cross-references: EMBL:AL031004
 A:Experimental source: cultivar Columbia; BAC clone F28M20

C;Genetics:
A:Map position: 4
A:Introns: 81/2; 146/3; 176/2; 205/1; 259/2; 300/3; 353/2; 381/3; 452/3; 473/2
A:Note: F28W20.140

Query Match	17.08;	Score	692.5;	DB	2;	Length	631;
Best local Similarity	34.2;	Pred.	No. 1.le-3;				
Matches	167;	Conservative	83;	Mismatches	194;	Indels	45;
Gaps	13;						
Qy	89	POKVLRFSEKIC--LKWOOTHRVAGAGLQNLGNTCFANAALQCLTYTTPPLANYMLSHSHSK	146				
Db	144	PREVLFFYEFVEYFNWDNELAPCGLMNCNSCFANVILQCLSWTPEPLVAYLLEKEGHR	203				
Qy	147	TCHAEAGFCMMCTQGAHTTQALSNEGVDVIKPMFVINEMRRIARHLRFGNQEDAHEFFLYQTV	206				
Db	204	ECMRNDWCFLECFQTHVERA--SQSRFPFPSPMNTISRLTNIIGTGLYGQGRQEDAHEFMRYAI	262				
Qy	207	DAMQAKACLN---GSNKLDRHUQTATTLVCQIFGGYILRSRVKCLNCKGVSDTFDPDPLDITLE	263				
Db	263	DMQSVCLDFBGGEKIVPPRSQETTLQYIFGGLQSQVQCTVCNHSDOYENNMMDLIVE	322				
Qy	264	1KA--AQSVNKALEGFVKPEOLDGNSVKCSKKQWPAASKRFTTHRSSNVLTLSLKRFPAN	322				
Db	323	MHGDAGLECLDQFTAEMLHGDNMYKDCRCSYVYKAKRLTIRRAPNLTIALKRYQOG	382				
Qy	323	FTGKGIAKDVKPYELDIRPYMSQPNGEPIYVYLYAVLVHTGF--NCHAGHYFCYIKASN	380				
Db	383	GRYGLKNKRISFPETLDINPYMSEGGDGDYVYLYAVIVHLDMLNASFFGHYICVKDFC	442				
Qy	381	GLWTQMDNSIVSTSDIRSVLSQQAYVIFY-----IRSH--DVKNGGELTHPTHS	427				
Db	443	GNWYRIIDDSEIESVELEDVLISQRAYMLLYSRIQARSSSSCLSRKSVQDEKKTDTTL-----	496				
Qy	428	PGQSSRPVLISQRYVVTNKQAAPGFIGPOLPSH-MKNPPHPLNGTGPKLQDTPSSSMS--SPN	485				
Db	497	DTESCVKELVESSMV-----GAIESRSSTHATIEDFVCBSQSPSP---SPSPSPSPSPS	546				
Qy	486	GNSSVNRASPNVNASASQNVNSVRNSVPIPEHPKKQKITISIHNNKLIPVRQCQSQPNLHNS	545				
Db	547	PSPSVLASECCSEVERDITLDSSENSSSIDSATDQHDQEVANGKDPVKYQA-----ADS	601				
Qy	546	LENFTKVPV	554				
Db	602	WSDPTGTPT	610				

RESULT 6

T05578
hypothetical protein F22K18.240 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
C:Accession: T05578
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hoheisel, J.; Mewes, H.W.; Ma
submitted to the Protein Sequence Database, February 1999
A:reference number: Z15419
A:Accession: T05578
A:Molecule type: DNA
A:Residues: 1-1008 <BEV>
A:Cross-references: EMBL:AL035356
A:Experimental source: cultivar Columbia; BAC clone F22K18
C:genetics:
A:Map position: 4
A:Introns: 94/2; 521/3; 550/2; 579/1; 633/2; 674/3; 727/2; 755/3; 825/3; 846/2
A:Note: F22K18.240

```

Query Match      15.4%; Score 626.5; DB 2; Length 1008;
Best Local Similarity 29.8%; Pred. No. 1.8e-29;
Matches 176; Conservative 96; Mismatches 218; Indels 101; Gaps 20;

QY      59  GAVVYSSSSVPDKS-----KPSQKQDQALGDGIAPPQKVLFFSEKICLKWQ 104
      ||||| :|||: |||||
      ||||| :|||: |||||

ph      479  GAVVQVHNSLHGRCGLKASVTVKVDWTRPKSENEMAGRHG-----HGKGLFYFVEVFKLY- 533
      ||||| :|||: |||||

```

105 QTHRV---GAGLQNLGNTCFANALQCLTYTPPLANVYMLSHHSKTCCHAEFGCMCTMOA 161
162 HITQALSNPGDVVKPMFVFNEMRIARHLRFGNQEDAHEFLQYTVDMQKACLNG-----216
593 LVRKAKEESP-LSPNGLLSLQNLGIFLNGKEEDAHEFLRVFVDTMQSVCIKASEYDM 651
217 --SNKLDHRTQATTLVCOIFGGYLRSRVKCLNCKGVSDTFDPYDLITLLEIKA-AQSUNKA 273
652 TKSSKL---BDTTLGLTGGYLRSKTKCMKQVKSLEKMDLTVEIDGDIISTLDDA 707
274 LEQVFKPEQLDGENSYKSCCKMVPASKRTIHRSSNVLTSLKRFANFANFGGKIADKVK 333
708 LRRTRFELDGENKYRCGCKSERAKKLIKITEPPNVLTIALKRFQAGFGKLNKLIR 767
334 YPEYLDIRPMSQNGEPVIVLYAVLVHTGF--NCHAGHYFCYIKASNGLWYQMDNSIV 391
768 FPETLDLAPYVSGGSEKSHDYKLYGVIVHLDVMAAFSGHYVCVIRNQN-KWYKADDSV 826
392 STDIRSVLSQOAVLVFIRSHDVKNKGELTHPTHSFGQSS--PRPVISQRVVTKQAAP 449
827 VTSVERILTKGAWLFAVCTPTPPRLAVCTKTEASNKGRVPLPKANKEKSTISRSVST 886
450 GFIFGQPLPSHMIKPPHNLGTGPKLDTFSS-----SMSSPNGNSVNRASPVNASAS 501
887 S--SPELSS---NTPGGRSGNIQSFYSSFORLQKILEEDSASDSSSLFDSNDECS 940
502 VQNSVN--RSSVTPHEPKKOKITISIHKLVPVQCSQPNLHNSLENTPKVPVPSSTIT 559
941 TDSTSMDDFADFIQGDH-----QGRAHQS-----ETPSPST-- 971
560 NSAVQSTSNASTMSVSSKVTPIPRSESCSQPVNMGSKLNSVLPVXGAE 610
972 -----SSSSSSSPFTRRSLR--SSPETYGTSRHO-----LPLGGE 1007

RESULT 7
T05075
hypothetical protein T6K21.70 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05075
R;Bevan, M.; Rieger, M.; Mueller-Auer, S.; Zipp, M.; Schaefer, M.; Bancroft, I.; Mewes,
submitted to the Protein Sequence Database, February 1998
A;Reference number: Z15397
A;Accession: T05075
A;Molecule type: DNA
A;Residues: 1-1082 <BEV>
A;Cross-references: EMBL:AL021889
A;Experimental source: cultivar Columbia; BAC clone T6K21
C;Genetics:
A;Map position: 4
A;Introns: 26/3; 66/2; 210/3; 302/3; 344/3; 378/3; 413/2; 428/3; 561/3; 602/1; 639/1; 69
A;Note: T6K21.70

Query Match
Best Local Similarity 12.4%; Score 504.5; DB 2; Length 1082;
Matches 185; Conservative 110; Mismatches 270; Indels 205; Gaps 25;

7 ASSESSDFAYQNPQSSEAVSPGMDAGSASGWSVSLNDVSNHTLSLGPVPGVAVYSSS 66
463 SSSSSDES--QSVSPSPINLDHDDDEQIPRNSTQALDDDD-----IWGDD 509

67 SVDPKSKP-SFQKQALG-----DGIAPQKVLFFSEKICLKWQTHRVGAGLQN 115
510 DLPTRETPWTPNVPSPGSGDDDDDDNDNSKNPERKSLFYGFR--QRPPEVTVGAGLWN 567
116 LGNTCFANALQCLTYTPPLANVYMLSHHSKTCCHAEFGCMCTMOAHITQALSNPGDV 173
568 LGNSCFNLNSVQCFTHVTPLESLLSPRYEVPCHGNEFCVTRAIRYHTEALRPERCP 627
174 IKMFVFNEMRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSKLNDRTHTQATTLVCQI 233

Db 628 IAPFFFDNLNYFSPDFQRYQOEDAHEFLQAFLEKLE---ICGSDRTSFRGDIISQ--DV 682
QY 234 FGGYLRGRVKCLNCKGVSDTFDPYDLITLLEIKAQASVNAKALEQFVKPQLDGENSYKSC 293
Db 683 FSGRLISGLRCNCNDYVSETYKSVGLSLEIEDVDLTLSALESTFRVEKLD--EQLTCDN 740
QY 294 CKKVPASKRTIHRSSNVLTSLKRFANFANFGG---KIAKDVKYPEYLDIRPMS--OP 347
Db 741 CNEKVSKEKQLLQKPLVATFHLKRFKN--NGLYMEKIYKHVKIPLFIDIQYMRNIQE 798
QY 348 NGEPIVYLYAVLVHTGFNCHAGHYFCYIKASNGLWYQMDNSIVSTSDIRSVLSQOAVL 407
Db 799 NEVSTKHLVALVEHFGYSVAYGHYSYVRSAPKLIWHFDDSKVTRIDEDMVLSDSYL 858
QY 408 PYIRSHDVKNKGELTHPTHSFGQSSPRPVISQRVVTKQAAPFGFIPQLPSHMIKPPH 467
Db 859 FYAR-----862
QY 468 NGTGPLKDTSPSSMSSPNGNSVNRASPVNASASVQWNSVNRSSVIPHPPKQKITISIH 527
Db 863 EGT-----RWFF-----SSVYEE-----874
QY 528 NKLPRQCSQPNLHNSLENTPKVPVPSSTITNSAVQSTSNASTMSVSS---KVTKPI 582
Db 875 -----MQPLVEASILNSPK-----SVLDSSTNGECLSEISYENGDKASKPC 916
QY 583 PRSESCSQPVNMGK---SKLNSVLPVYAGABSSDESDSESKGLKENGIGTIV-----S 633
Db 917 DSAGVCNQHVKTCKDFVSLGNDVFL--SAESSS-----GEESPMLGELLPLDPDD 965
QY 634 SHSPGQDAED-----EEATPHELOEPMTLNGANSADSDPKENGLAPDAGSCQGPAL 687
Db 966 SYSCTEKESDSCIAIRATIRDFPPLLD-QNQESTSPKLOERTFEMQLIQMBETT 1024
QY 688 HSENPFAK-----ANGLPGKLMAPALLSLPEDKILETFRLSNKLKGS 729
Db 1025 KSQEPWKPQLSSINIAADSMEAEFVYGLMKKPSPRARELLDQAISTNGS 1074

RESULT 8
S57591
hypothetical protein YMR223w - Yeast (Saccharomyces cerevisiae)
N;Alternate names: hypothetical protein YMR9959.05
C;Species: Saccharomyces cerevisiae
C;Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 19-Apr-2002
C;Accession: S57591
R;Skellton, J.; Churcher, C.M.
submitted to the EMBL Data Library, June 1995
A;Reference number: S57587
A;Accession: S57591
A;Molecule type: DNA
A;Residues: 1-471 <SKE>
A;Cross-references: EMBL:Z49393; NID:g887599; PIDN:CAA90194.1; PID:g887604; GSPDB:GN0001;
A;Experimental source: strain AB972
C;Genetics:
A;Gene: SGD:URP8; MIPS:YMR223w
A;Cross-references: SGD:S0004836
A;Map position: 13R

Query Match
Best Local Similarity 10.5%; Score 428; DB 2; Length 471;
Matches 104; Conservative 68; Mismatches 125; Indels 34; Gaps 11;

QY 111 AGLQNLGNTCFANALQCLTYTPPLANVYMLSHHSKTC--HAEGFCMCTMQ-----160
Db 137 SGLINMGSTCFMSSILQCLIHNFYFIRHSNQSIHNSNCKVRSPPCKFSCALDKIVHLYG 196
QY 161 -AHITQALSNPGDVVKP--MFVINEMRIARHLRFGNQEDAHEFLQYTVDMQKAC---213
Db 197 ALMTKQASSSTSTNRTGTGFIYLLTCAWLNQNLAGYSQDDAHEFWOIFINHQSVLD 256
QY 214 LNSGNKLDR--HTQATTLVCOIFGGYLRSRVKCLNCKGVSDTFDPYDLITLLEIKAQSV 270

RESULT 10
T37611
ubiquitin carboxyl-terminal hydrolase - fission yeast (Schizosaccharomyces pombe)

Db 283 NTVDEYVDIE-----YSLFAVVHVHVGSPNGHGVSLVKSHNH-WLFFDDSEV 329

QY 389 SIVSTSDIRSVL-----SQQAYVLFY 409

Db 330 ELIEBASVQTFPGSSQEQVSSNTDGHYILLY 359

RESULT 14

S50277

ubiquitin-specific proteinase UBP5 (EC 3.4.-.-) - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YER144c

C:Species: Saccharomyces cerevisiae

C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 29-Sep-1999

C:Accession: S50277; S50647

R:Xiao, W.; Fontanie, T.; Tang, M.

Yeast 10, 1497-1502, 1994

A:Title: UBP5 encodes a putative yeast ubiquitin-specific protease that is related to th

A:Reference number: S50277; PMID:95176708; PMID:781889

A:Accession: S50277

A:Molecule type: DNA

A:Residues: 1-805 <XIA>

A:Cross-references: EMBL:U10082; NID:g5953376; PIDN:AAC48928.1; PID:g5953377

R:Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A:Description: The sequence of S. cerevisiae cosmids 8229, 9115, 9132, 9981, and lambda

A:Reference number: S50647

A:Accession: S50647

A:Molecule type: DNA

A:Residues: 1-805 <DIE>

A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64671.1; PID:g603384; MIPS:YER144c

C:Genetics:

A:Gene: SGD:UBP5

A:Cross-references: SGD:S0000946; MIPS:YER144c

A:Map position: 5R

C:Superfamily: deubiquinating enzyme SSV7

C:Keywords: hydrolase

Query Match 8.2%; Score 332; DB 2; Length 805;

Best Local Similarity 26.1%; Pred. No. 4-5e-12;

Matches 123; Conservative 75; Mismatches 162; Indels 112; Gaps 22;

QY 8 SESSDFSAVQNGSSEAVSPGMDAGSANGAVSLNVSNTLSLGPVPGAVVSSSS 67

Db 376 SKLNTPTSTQNKANTVERISP-DIRAAQA-----HAY-----LPPASNVFSR 417

QY 68 VPKSKSPQKQALGDGIAPPQKVLFPSEKI CLKWQOQTHRVGAGLQNLGNTCFANAALQ 127

Db 418 IP-----PLPOONLS-----SSRQTLNNSQVLLD-----LVGLENI GNCVMNCILQ 462

QY 128 CLAVTTPPLANYMLSHR-----SKTCHAEAGFCMCTMOAHITQALSNPGDVIKP 176

Db 463 CLVGTDLVRMFLDNTYLNFINFDSRSGSKGLLAKNFALLVN-NMHRGAFTPPNVRTIP 521

QY 177 MFVINEMRRIARHLR-----FGNQEDAHEFLQYTVDMQKAC-LNGSNKLDH----- 223

Db 522 VQTI-QFKKICGHINPMYSMDQDCQFCQFLDGLDHEDLNQNGSKKHLKQLSDBERM 580

QY 224 -----TQATTL-----VCOIFGGLRSRVKCLNCKGVSDTFDPLDITILEIK 265

Db 581 REKMSIRKASALEWERILLTFSAILIDLFQOQYASRIQCQVCEHTSTITYQTFVSLVSPVP 640

QY 266 AAQSVN--KALEQFVKPQLDGENSYKSKCKKMPVASKRFTIHRSSNVLTSLSKRPFANF 323

Db 641 RVKTCNILDCEFTFKCERLGVDEQWSCPKLKKQPSKQLKITRLPKKLIINLKRFDN- 699

QY 324 TGGKIADK---VKPEYLDIRPYMSQP-NGEPIV-----VLYAVLVHTGF 365

Db 700 ---QMNKNVQVPEYSLDTPYWARDFNHEIAVNEIDIPTRGQVPPFIRLYGVACHSG- 755

QY 366 NCHAGHYFCYI-KASNGLWYQNDSI-----VSTSDIRSVLSQQAYVLFYIR 411

Db 756 SLYGGHYTSYVYKPKKGMWFFDDSLYRPITFT-----EFTPSAYVLFYER 803

RESULT 15

T06097

ubiquitin-specific proteinase (EC 3.4.-.-) UBP3 - Arabidopsis thaliana

N:Alternate names: protein T5J17.80

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06097

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184

A:Accession: T06097

A:Molecule type: DNA

A:Residues: 1-371 <BEV>

A:Cross-references: EMBL:AL035708; GSPDB:GN00062; ATSP:T5J17.80

A:Experimental source: cultivar Columbia; BAC clone T5J17

C:Genetics:

A:Gene: ATSP:T5J17.80; UBP3

A:Map position: 4

A:Introns: 39/3; 78/3; 111/3; 177/3; 238/2

C:Keywords: hydrolase

Query Match 7.9%; Score 323.5; DB 2; Length 371;

Best Local Similarity 25.4%; Pred. No. 5.1e-12;

Matches 99; Conservative 56; Mismatches 143; Indels 91; Gaps 12;

QY 79 DOALGDGIAPPQKVLFPSEKI CLKWQOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANY 138

Db 9 EKALGDQFPGEGRYF-----GFENFGNTCYCNVQLQALYFCVFPREQ 50

QY 139 MLSH-----EHSKTCHAEAGFCMCTMOAHITQALSNPGDVIKPMFINEMRRIA 187

Db 51 LLEYTTSNKSVDADAEENLMTCLADLF-----SQISSQKKKTG-VIAPKRFVQRLKKQN 102

QY 188 RHLRFGNQEDAHEFLQY-----TVDMQKAC-----LNG-----SNKL 220

Db 103 ELFRSYMHQDAHEFLNLLNEVVDILEKEAKATKTEHTSSSSSSPEKIANGLVQPVQANGV 162

QY 221 DRHTQATTLVCOIFGGLRSRVKCLNCKGVSDTFDPLDITILEIKAAQSVNKALEQFVKP 280

Db 163 VHKEPIVTWVHNIFQGIILTNETRCLRCETVTARDETFLDLSLDIEQNSSITSCIKNFSST 222

QY 281 EQLDGENSYKSKCKKMPVASKRFTIHRSSNVLTSLSKRPFANFTG-----GKIARDVKYPE 336

Db 223 ETLHAEDKFFCDKCCSLQEAQKMKIKKPHILVIHLKRFKYIEQLGRYKLSYRVVFEPL 282

QY 337 YL-----DIRPYMSQPNGEPIVYVLYAVLVHTGNCHAGHYFCYIKASNGLWYQNDSIV- 391

Db 283 ELKLSNTEVEPYAD-----VEYSILFAVVHVHVGSPNGHGVSLVKSHNH-WLFFDDENVE 335

QY 392 -----STSDIRSVLSQQAYVLFY 409

Db 336 MIEESAVQTFPGSSQEQVSSNTDGHYILLY 364

Search completed: August 10, 2004, 20:00:19

Job time : 21 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: August 10, 2004, 19:55:58 ; Search time 43 Seconds
(without alignments)
5686.665 Million cell updates/sec

Title: US-10-049-745-4
Perfect score: 4070
Sequence: 1 MTIVDKASESDPSAYQNP.....AAESLEEDPAAASLRPFSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rviris.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1808.5	44.4	1016	13 Q802X0	Q802X0 brachydanio
2	1375	33.8	548	4 Q9H9C5	Q9H9C5 homo sapien
3	915	22.5	919	5 Q86U16	Q86U16 dictyosteli
4	895	22.0	466	11 Q8BTN5	Q8BTN5 mus musculu
5	894	22.0	545	11 Q923V2	Q923V2 mus musculu
6	890.5	21.9	545	11 Q923V3	Q923V3 mus musculu
7	889.5	21.9	545	11 Q55191	Q55191 mus musculu
8	888	21.8	545	11 Q55190	Q55190 mus musculu
9	763.5	18.8	842	10 Q84P80	Q84P80 mus musculu
10	760	18.7	457	3 Q74442	Q74442 oryza sativ
11	748	18.4	859	10 Q9FPS4	Q9FPS4 schizosacch
12	737	18.1	661	10 Q9FPS2	Q9FPS2 arabidopsis
13	735.5	18.1	1038	5 Q8IQ60	Q8IQ60 drosophila
14	735.5	18.1	1085	5 Q9VRP5	Q9VRP5 drosophila
15	734	18.0	661	10 Q8VZF5	Q8VZF5 arabidopsis
16	733.5	18.0	948	5 Q8IQ58	Q8IQ58 drosophila

17	733.5	18.0	968	5 Q960Q4	Q960Q4 drosophila
18	733.5	18.0	995	5 Q8IQ59	Q8IQ59 drosophila
19	733.5	18.0	1015	5 Q8IQ57	Q8IQ57 drosophila
20	730.5	17.9	1038	5 Q86NM9	Q86NM9 drosophila
21	725.5	17.8	948	5 Q7YTX7	Q7YTX7 drosophila
22	720	17.7	891	10 Q9SHG9	Q9SHG9 arabidopsis
23	704	17.3	672	10 Q9SJA1	Q9SJA1 arabidopsis
24	692.5	17.0	631	10 Q81780	Q81780 arabidopsis
25	686.5	16.9	731	10 Q9FKP5	Q9FKP5 arabidopsis
26	686.5	16.9	924	10 Q9FPS9	Q9FPS9 arabidopsis
27	646.5	15.9	632	10 Q9LJK9	Q9LJK9 arabidopsis
28	645	15.8	829	10 Q9FUL7	Q9FUL7 arabidopsis
29	626.5	15.4	1008	10 Q9FPS8	Q9FPS8 arabidopsis
30	626.5	15.4	1008	10 Q9SB51	Q9SB51 arabidopsis
31	555.5	13.6	285	4 Q8IXW9	Q8IXW9 homo sapien
32	555.5	13.6	732	10 Q9FIQ1	Q9FIQ1 arabidopsis
33	555.5	13.6	737	10 Q9PPS6	Q9PPS6 arabidopsis
34	504.5	12.4	695	10 Q9FPS7	Q9FPS7 arabidopsis
35	504.5	12.4	1082	10 Q49688	Q49688 arabidopsis
36	437	10.7	703	5 Q95TK9	Q95TK9 drosophila
37	437	10.7	735	5 Q9VUQ9	Q9VUQ9 drosophila
38	437	10.7	735	5 Q9VVR1	Q9VVR1 drosophila
39	420	10.3	520	4 Q8WVD0	Q8WVD0 homo sapien
40	415.5	10.2	525	11 Q9JIG5	Q9JIG5 mus musculu
41	413	10.1	938	5 Q8IQ27	Q8IQ27 drosophila
42	411	10.1	856	5 Q9VR54	Q9VR54 drosophila
43	403	9.9	337	5 Q8IQ26	Q8IQ26 drosophila
44	402.5	9.9	540	11 Q8BSW2	Q8BSW2 mus musculu
45	393	9.7	896	5 Q8MQX4	Q8MQX4 drosophila

ALIGNMENTS

RESULT 1

Q802X0	PRELIMINARY;	PRT; 1016 AA.
ID Q802X0		
AC Q802X0;		
DT 01-JUN-2003 (TREMBLrel. 24, Created)		
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)		
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE Similar to KIAA1453 protein (Fragment).		
OS Brachydanio rerio (Zebrafish) (Danio rerio).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
OC Cyprinidae; Danio.		
OX NCBI_TaxID=7955;		
RN [1]		
RP SEQUENCE FROM N.A.		
RC STRAIN=AB;		
RL Strausberg R.;		
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR EMBL; BC047168; AAH47168.1; -		
DR GO; GO:0004197; P:cysteine-type endopeptidase activity; IEA.		
DR GO; GO:0004379; P:glycylpeptide N-tetradecanoyltransferase activity; IEA.		
DR GO; GO:0004221; P:ubiquitin thiolesterase activity; IEA.		
DR GO; GO:0006499; P:N-terminal protein myristoylation; IEA.		
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.		
DR InterPro; IPR003006; Iq_MHC.		
DR InterPro; IPR000903; Nmt.		
DR InterPro; IPR001394; Peptidase_C19.		
DR Pfam; PF00443; UCH; 1.		
DR PROSITE; PS00290; IG_MHC; 1.		
DR PROSITE; PS00976; NMT_2; 1.		
DR PROSITE; PS00972; UCH_2_1; 1.		
DR PROSITE; PS00973; UCH_2_2; 1.		
DR PROSITE; PS00973; UCH_2_3; 1.		
DR PROSITE; PS0235; UCH_2_3; 1.		
FT NON TER 1016 1016		
SQ SEQUENCE 1016 AA; 113138 MW; 05AFD86206CC9DAA CRC64;		

Query Match 44.4%; Score 1808.5; DB 13; Length 1016;
Best Local Similarity 50.9%; Pred. No. 6.3e-120;
Matches 390; Conservative 96; Mismatches 199; Indels 81; Gaps 20;

```

RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK022913; BAB14305.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR003006; Ig MHC.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH; 1.
DR PROSITE; PS00290; IG_MHC; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW Hypothetical protein.
SQ
SEQUENCE 548 AA; 60332 MW; 2CSED382105AC2C6 CRC64;

Query Match 33.8%; Score 1375; DB 4; Length 548;
Best Local Similarity 47.3%; Pred. No. 1.9e-89;
Matches 286; Conservative 81; Mismatches 152; Indels 86; Gaps 12;

QY 1 MTIVDKASESSDPEAYONOPGSEAVSPGDMAGSASGAVSSLDV----- 47
Db 1 MPIVDKLKEAL-----KPERKDSADDGELGKLASSAKVLLQKIEFEPASKSFYQL 53
QY 48 -----SNHTLSLGPVPGAVVY-SSSVDPKSPQKQDQALGDIAPQKVLFPSEKICLK 102
Db 54 EALKSVYLLNPKTETGASRHKSGDDPPARRRGSEHTYESCGDGVGPAPQKVLFPETERLSLR 113
QY 103 WQOQTHRVGAGLONIGNTCFANALQCLTYTPPLANYMLSHSEKTCCHAEFGCMCTMQAH 162
Db 114 WERVFRVAGLHNLGNTCFNLATIQCLTYTPPLANYLLSEHARSCHQGSFCLCMQNH 173
QY 163 ITQALSNPGDIVKPMFVINEMRRIARHFRGNOEDAHEFLQYTVDMQKACLNGSNKLD 222
Db 174 IVQAPANGNAIKPVSEFIRDLKLIARHFRGNOEDAHEFLRYTIDAMQKACLNGCAKLD 233
QY 223 HTQATTLVCOIFGYSRLSRVKLCKGVSOTDFPDYDITLLEIKAAQSVNKALEOFVKPEQ 282
Db 234 QTQATTLVHQIFGYSRLSRVKCSVSDTYDPLDVALEIRQAAIVRALELUFKADV 293
QY 283 LDGENSVKCKCKQWPAKRFTHRSNNVLTSLKRFANFTGGKIAKDKVKPYEYLDIRP 342
Db 294 LSGENAYMCAKCKKVPASKRFTIHRSSNVLTSLKRFANFSGGKIYKDVGYPEFLNIRP 353
QY 343 YMSQNGPEPTVYLYAVLVHTGFNCHAGHYFCYTKASNLWYQMDSVTSDTSRVSLSQ 402
Db 354 YMSQNGDPMYGLYAVLVHSGYCHAGHYCYVKASNGQWYQMDSLVHSSNVKVLNQ 413
QY 403 QAYVLYFIRSHDVKNKGELTHPTSPG-QSSPRPVISORVVTNKQAARGFTGPQ-LPSHM 460
Db 414 QAYVLYFIR-----IPGSKSPGELISR---TGSSTLFG--RPSVIPDHS 453
QY 461 IKNPPLHNGTQPL--KDTSPSSMSSNGSNVNRASPVNASASVQNVNRSSVPIPEHPK 518
Db 454 KKNIGNGIISPLTKGKQDQSGTKMKPHTEI--GVPIRNGSTLGLKSQNGCIPPKLPS 511
QY 519 KOKITISHNKLPVRQCSQPNLHNSLENPTKPVPSSTIINSAVQSTSNASTMVSVKY 578
Db 512 GS-----FSPKLSQTPTHTPTLDDPGK-----KV 536
QY 579 TKPIP 583
Db 537 KKPA 541

RESULT 3
Q86J16
ID Q86J16 PRELIMINARY; PRT; 919 AA.
AC Q86J16;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.

```


RESULT 6		
Q923V2		
ID	Q923V2	PRELIMINARY; PRT; 545 AA.
AC	Q923V2;	
DT	01-DEC-2001 (TREMBlrel. 19, Created)	
DT	01-DEC-2001 (TREMBlrel. 19, Last sequence update)	
DT	01-OCT-2003 (TREMBlrel. 25, Last annotation update)	
DE	Deubiquitinating enzyme 2A.	
GN	DUB2 OR DUB2A.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
OX	NCBI_TaxID=10090;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=129/W;	
RA	Baek K.-H., Mondoux M.A., Jaster R., Fire-Levin E., D'Andrea A.D.;	
RT	"DUB-2A, a new member of the DUB subfamily of hematopoietic	
RT	deubiquitinating enzymes.";	
RL	Blood 98:0-0(2001).	
DR	EMBL; AF393638; AAK77003.1; --.	
DR	MGI; 107698; Dub2.	
DR	GO; GO:0004843; F:ubiquitin-specific protease activity; IDA.	
DR	GO; GO:0016579; P:protein deubiquitination; IDA.	
DR	InterPro; IPR001394; Peptidase_C19.	
DR	Pfam; PF00443; UCH; 1.	
DR	PROSITE; PS00972; UCH 2.1; 1.	
DR	PROSITE; PS00973; UCH 2.2; 1.	
DR	PROSITE; PS0235; UCH 2.3; 1.	
SQ	SEQUENCE 545 AA; 2FFDF097313FB88D CRC64;	
Query Match 22.0%; Score 894; DB 11; Length 545;		
Best Local Similarity 36.3%; Pred. No. 3.3e-55;		
Matches 208; Conservative 96; Mismatches 193; Indels 76; Gaps 14;		
QY	69 PDKSPSPQ---KDQALGDGIAPPQKV--LFFSEKICLKQWQTHRVGAGLQNLGNTCFAN 123	
DB	12 PALSSPGAQLHQDEA-----QVVVELTANDKPSLSEWCPCPGCGLQNTGNSCYLN 63	
QY	124 AALQCLTYTPPLAN YMLSHSHTCHAEFCMMCTQWAIHQAL--SNFGDVTKPMFVIN 181	
DB	64 AALQCLTHTPPLADYMLSQEYSQTCSPGCKMCAMEAHVTQSLHSHSGDVNPKSQILT 123	
QY	182 EMRRIARHLRFGNOEDAHFLOQTVDMOKACILNGSNKLDHRTQATTLVCQIFGGYLRSR 241	
DB	124 S-----AFHXH--QOEDAHEFLMPTLETMHESCLQVHRQSEPTSEDSPPHDI FGGLRWSQ 177	
QY	242 VKCLNCKGVSDTDPDYLDITLTKAAQSVNKALEQFVKPEQLDGENSYKSKCKKMPAS 301	
DB	178 IKLHCQGTSDTYDRFLDPLDISSAQSVNQALWDTEKSEELRGENAYYCGRCRQKMPAS 237	
QY	302 KRFTIHRSSNVLTSLKRFANFTGGKIADVKYPEYLDTRPYMSQPNQGEPIVIVLYAVLV 361	
DB	238 KTLHIHSA PKVLLLVLRFSASMGKLDKRVSPYEFDLKPYLSQPTGGPLPYALYAVLV 297	
QY	362 HTGFNCHAGHYFCYIKASGLWYQMDNSIVSTSDIRSVLSQQAAYLVFIYIRSHDVXNGGEL 421	
DB	298 HEGATCHSHGYFSYVKGAGHKYKMDDTKVTSCDVTSLNENAYLVFYVQOTDLKE----- 353	
QY	422 THPTSPGQSSPRPVISQRYVTNKAAPGFIFGQLPS---HMKINPPH--LNGTGPKLKDTP 477	
DB	354 -----VSIDMPEGRITHEVLDPYQLKRRKKHKK 383	
QY	478 SSSMSPNGNSVNRRASPVNASASVQWNSVNRSSVPEHKKQKITIISHNKLPVQCOS 537	
DB	384 KSPCTEDVGSPSKNR-----EKKATKETSIGEGVKLQEKHKKAGQKHENTKL--VPOQN 437	
QY	538 QPNLHNSLENPTKPVPS--TITNSAVQSTSNASTMSVSKVTPIPRS-----ES 587	
DB	438 HQKLQKHRRNEILPOFNQHKQTCQSLNTEGELDLPAIDAIVIH--LFRSIANWRGDTFDK 496	
QY	588 CSQFVWNGKSLNSSLVLPYGAESSEDSDESK 620	
DB	497 VNPQWNNADRLITSQDLVNTGLLCRQEGRRRSK 529	

Query Match 21.9%; Score 890.5; DB 11; Length 545;		
Best Local Similarity 35.5%; Pred. No. 5.8e-55;		
Matches 210; Conservative 91; Mismatches 191; Indels 99; Gaps 14;		
QY	69 PDKSPSPQ---KIQALGDGIAPPQKV--LFFSEKICLKQWQTHRVGAGLQNLGNTCFAN 123	
DB	12 PALSSPGAQLHQDEA-----QVVVELTANDKPSLSEWCPCPGCGLQNTGNSCYLN 63	
QY	124 AALQCLTYTPPLAN YMLSHSHTCHAEFCMMCTQWAIHQAL--SNFGDVTKPMFVIN 181	
DB	64 AALQCLTHTPPLADYMLSQEYSQTCSPGCKMCAMEAHVTQSLHSHSGDVNPKSQILT 123	
QY	182 EMRRIARHLRFGNOEDAHFLOQTVDMOKACILNGSNKLDHRTQATTLVCQIFGGYLRSR 241	
DB	124 S-----AFHXH--QOEDAHEFLMPTLETMHESCLQVHRQSEPTSEDSPPHDI FGGLRWSQ 177	
QY	242 VKCLNCKGVSDTDPDYLDITLTKAAQSVNKALEQFVKPEQLDGENSYKSKCKKMPAS 301	
DB	178 IKLHCQGTSDTYDRFLDPLDISSAQSVNQALWDTEKSEELRGENAYYCGRCRQKMPAS 237	
QY	302 KRFTIHRSSNVLTSLKRFANFTGGKIADVKYPEYLDTRPYMSQPNQGEPIVIVLYAVLV 361	
DB	238 KTLHIHSA PKVLLLVLRFSASMGKLDKRVSPYEFDLKPYLSQPTGGPLPYALYAVLV 297	
QY	362 HTGFNCHAGHYFCYIKASGLWYQMDNSIVSTSDIRSVLSQQAAYLVFIYIRSHDVXNGGEL 421	
DB	298 HEGATCHSHGYFSYVKGAGHKYKMDDTKVTSCDVTSLNENAYLVFYVQOTDLKE----- 353	
QY	422 THPTSPGQSSPRPVISQRYVTNKAAPGFIFGQLPSHMKINPPH--LNGTGPKLKDTPSS 480	
DB	354 -----CSIDMPEGRITHEVLDPYQLKRRKKHKKSP 386	
QY	481 MSSPNGNSVNRRASPVNASASVQWNSV-----NRSVPEHKKQKITIISIH 527	
DB	387 CTEDVGSPSKNRKIKATKETSIGEGVKLQEKHKKAGQKHENTKLVPQEQNHOKLGOKHR 446	
QY	528 NK--LPVRCQSQNLHNSLENPTKVPSPSTITNSAVQSTSNASTMSVSKVTPIPRS 585	
DB	447 NNEILPOEQ-----NHQKTCQSLNTEGELDSPADA-----IVIHLPSS 485	


```

354 -----VSIDMPGRVHEVLDPEYQLKSRKKHKK 383
Db

478 SSSMSSPENGSSVNRASVPNASASVQNSV-----NRSSVIPHPKKQKI-- 522
QY

384 KSPCTEDAGFCCKREKREATKETSILGEGKYAQEKXHKAGQKHENTKLVPOEQNHQKLGQ 443
Db

523 TISHNKLPRVQCOSQPNLHNSLENP--TKVPSSSTITNSAVOSTSNASTMSVSKYTK 580
QY

444 KHRINEILPQNHOKA---QGSRLNTEGELDPADAIVIHLLRSTENWG-----R 491
Db

581 PIPRSESCSQPMNGSKSLNSVLVPYGAESSEDSDESK 620
QY

492 DAPDKE--NQPWNADRLTTSQDPVNTGQLCRQGRRRSK 529
Db

RESULT 9
ID Q84P80 PRELIMINARY; PRT; 842 AA.
AC Q84P80;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Purative ubiquitin-specific protease (Fragment).
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Euharoidae; Oryzae; Oryza.
NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=30843;
RX MEDLINE=22584437; PubMed=12684538;
RA Cooper B., Clarke J., Budworth P., Kreps J., Hutchison D., Park S.,
RA Guimil S., Dunn M., Luginbuhl P., Ellero C., Goff S.A., Glazebrook J.;
RT "A network of rice genes associated with stress response and seed
RT development."
RL Proc. Natl. Acad. Sci. U.S.A. 100:4945-4950(2003).
DR EMBL; AY224488; AAC72607.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
DR Protease.
FT NON_TER
SQ SEQUENCE 842 AA; 93153 MW; F3BFC7AADE5E93620 CRC64;

Query Match 18.8%; Score 763.5; DB 10; Length 842;
Best Local Similarity 30.5%; Pred. No. 1.2e-45;
Matches 196; Conservative 117; Mismatches 239; Indels 91; Gaps 17;

QY 110 GAGLQNLGNTCPANAALQCLTYTPPLANVLSHESKTCBAGFCMCTMQAHIITCALSN 169
Db 1 GAGLQNLGNTCPANAALQCLTYTPPLANVLSHESKTCBAGFCMCTMQAHIITCALSN 169
QY 170 PGDVTIKPMFVINEMRIRARHLRFGNOEDAHEFLQVTDAMQACL-NGSNKLDLDRHTQAT 228
Db 61 TGKIYTPSQIVKVLKICISFRNSQEDAHALVNLLESMHKCLPSGVPSSESPAYEKS 120
QY 229 LVQCITFGYLRSRVKCLNKGVSDFDPYLDITLBIKAAQSVNKALEQVQPEQLD-GEN 287
Db 121 LVHKIFGRLRSQVKTCQCHGCKNFKDPLDLSLDIGKATSLVRALQNLFTABELLDGGEK 180
QY 288 SYKCKSKKMWPKASKEFTTHRSNVLTLTLKRPANFT-GGKIADQVKYPEYLDIRPYMSQ 346
Db 181 QYQORCRKRVKAKRFTIDKAPYVLTILHKRFPNPREKIDKKVDFOFMDLDFITSD 240
QY 347 PNGETIVVYLVAVLHTGTFCHAGYFCYIKASGLWYQMNDISIVSTSDIRSVLSQAYV 406

```

```

241 SKVSNLYSLYGLVHAGWNTQSGHYCYFVTSSGWHNLDNQVRQVREADVLRQKAYM 300
QY

407 LFYIRSHDVKNGBELTHPTSPGQSSRRP-----VISQVVVTKQAAPGFIGPOLPSHM 460
Db

301 LFYVRDR-----VGNPTPRKDNITANMPARRTIPEKISG---LSDMIQSGV 343
QY

461 IKNPPLHNGTGFLKUTPSSSMSSPENGSSVNRASVPNASASVQNSVNRSSVIPHPKKQ 520
Db

344 IE--AKLNGS-----SSPYGDKRLHIGNGNSI-----KTSREHYLKKDGKTE 384
QY

521 KITSIHNLKPVRCQSQPNLHNSLENP--TKVPSSSTITNSAVOSTSNASTMSVSKYTK 580
Db

385 APKASENNGLASTQASAPDIDGATLSAQSQKITSTGHREYSSSDRSASLTHIVNQAVA 444
QY

581 PIPRSESCSQPMNGSKSLNSVLVPYGAESSEDSDEES--KGLGHENGIGTIVS 633
Db

445 MVPSQE--LQPKVDGLTDTSTSLGNGNAILSRNKQTSQHQNPFSPASHGDKDTGAG-LAA 501
QY

634 SHSPQDAEDEATFHELOEPMTLGA--NSADSDSDPKENGLADPGAGCCQOPALHSEN 691
Db

502 QTFPTKDA-----IVSNGVVPSSRDPISSEKVCGL-----QKSIKQDD 539
QY

692 PFAKANGLPGLKMPAPLLSLPEDKILETFRLSNKLKSTDEMS 734
Db

540 KTVKE-----LPISENNIVSGLERVNARKQTSSEVS 570
QY

RESULT 10
ID Q74442 PRELIMINARY; PRT; 457 AA.
AC Q74442;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase.
GN SPC1682.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h;
RA Murphy L., Harris D., Lyne M., Rajandream M.A., Barrell B.G.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DDJB databases.
DR EMBL; AL031525; CAA20679.1; -.
DR PIR; T41069; T41069.
DR GeneDB SPombe; SPC1682.12c; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0004678; F:hydrolase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro; IPR001394; Peptidase_C19.
DR Pfam; PF00443; UCH_1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
DR Hydrolase.
KW Hydrolase.
SQ SEQUENCE 457 AA; 51581 MW; 56AAD2AA996BFF3 CRC64;

Query Match 18.7%; Score 760; DB 3; Length 457;
Best Local Similarity 38.8%; Pred. No. 8.9e-46;
Matches 158; Conservative 74; Mismatches 143; Indels 32; Gaps 8;

QY 28 PGMDAGSAGSAGVSLNDVSNHTLSLGPVGAVVYSSSSVDPKSPQKQ----- 80
Db 31 PAWSSPSVPEGTIVLNNPKQSTVS-----RKSFSAFTSPTNRKSGSSQYKSSG 84
QY 81 ALGDG-----IAPQKVLFPSEKICLKWQOQTHRVGAGLQNLGNTCFANAALQCLT 130
Db 85 TRGDGANDFVDEDPAFIPPARILFPEEKUSMEWDNIMFNAPGLVNLGNTCFMNSVLQMT 144

```

Db	242	EQSCHCSNKFDPFLDLSLDISKVDSLQRLGRFTAVELLDNGAKVYQCERCKQKVKAKQ	301
Qy	304	FTIHRSSNVLLSLIKRFANFTGGKIADKVPYELDIRPYMSQNGEPIVYVLYAVLVHT	363
Db	302	LTVSKAPVVLTVHLKRFEARSEKIDRKVDFTSAIDMKRPFVSGPHEGNIKLYTLXGLVHY	361
Qy	364	GFNCCHAGHYFCYIKASNGLWQMDSIVSTDIRSVLSQAYVLPIYIRSHDVNGGELTH	423
Db	362	GRSSHGHYACFVRTSGMWYSLDDNRVQVSEKTVFNQKAYMLFYVR--DRQNA----	414
Qy	424	PTHSPGOSPPRVISQSRVVTNKQAPGFIGPQLPSHMIKNPPLHNGTGLKDTSPSSMSS	483
Db	415	---VPKNSVP-VVKESFATNRSL-----IVASNIKO--QVNGSTVIKECGFGALVA	461
Qy	484	PNGNSVNRASPPVNASASVQ---NWSVNRSSVIEPHPKKQKITIISHNKLVPVRCQ---	536
Db	462	-NGLAPLKS CGSP TPAVLTKDLNAKETQNNAI--SNVEAKEILETENGSAFVKTCDLAAP	519
Qy	537	---SOPNLHNSLENPKPKPPSPSITITNSAVOSTNSATMSVSSVKVTKPIPRSESCSPVM	593
Db	520	TVLVQKDLNTKEIPQKRVPLPQANGEGSLVKEDSKAACLILPEKVS-----PHL	568
Qy	594	NGSKLNSSVLVPYGAESSDSDESGLGKENG-----IGTIVSHSPGQDAEDEEA	646
Db	569	DGSNAAQTLVKLPYLGPKANSVEEKSNLNLNPANSLKVINVSNGPPVEKAVLIDQT	628
Qy	647	TPHELOEPMT-----LNGANSADSDSPKENG LAPDGA SCQPALHGHENPFKANGLP GK	702
Db	629	MGHLEBSATSIESLKLTSERLITTK-----NTRPEPTKYLKVEFK	671
Qy	703	LMFAPLILSLPDKILETFRLSNKLKG	728
Db	672	FFKL-ALGLRKKVQRRERLSTTVAG	696

RESULT 12

Q9FPS2

ID

Q9FPS2

PRELIMINARY;

PRT;

661 AA.

AC

Q9FPS2;

DT

01-MAR-2001 (T-EMBLrel. 16, Created)

DT

01-MAR-2001 (T-EMBLrel. 16, Last sequence update)

DT

01-OCT-2003 (T-EMBLrel. 25, Last annotation update)

DE

Ubiquitin-specific protease 25.

GN

UBP25.

OS

Arabidopsis thaliana (Mouse-ear cress).

OC

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC

eurosid II; Brassicales; Brassicaceae; Arabidopsis.

NCBI_TaxID=3702;

[1]

SEQUENCE FROM N.A.

RP

MEDLINE=20567829; PubMed=11115897;

RX

Yan N., Doelling J.H., Falbel T.G., Durski A.M., Vierstra R.D.;

RA

"The Ubiquitin-Specific Protease Family from Arabidopsis. AtUBP1 and 2

RT

Are Required for the Resistance to the Amino Acid Analog Canavanine.";

RL

Plant Physiol. 124:1828-1843 (2000).

DR

EMBL; AF302673; AAG42763.1; -.

DR

GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

DR

GO; GO:0008233; F:peptidase activity; IEA.

DR

GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.

DR

GO; GO:0008511; P:ubiquitin-dependent protein catabolism; IEA.

DR

InterPro; IPR001394; Peptide_C19.

DR

Pfam; PF00443; UCH_1.

DR

PROSITE; PS00972; UCH_2.1; 1.

DR

PROSITE; PS00973; UCH_2.2; 1.

DR

PROSITE; PS00235; UCH_2.3; 1.

DR

Protease.

KW

SEQUENCE

661 AA; 73293 MW; D47DEAD5E8D9B8E CRC64;

Query Match

18.1%; Score 737; DB 10; Length 661;

Best Local Similarity

29.8%; Pred. No. 6.5e-44;

Matches

200; Conservative 113; Mismatches 230; Indels 128; Gaps 21;

Db 72 SVLENLKSXIVIKPGNP-----GAINGFSGKNN-----TGKLVGANGHDNNGARKQAE 120
 Qy 74 -PSPQDQ-----ALGDGIAPQKVLFPSEKICLKWOQTHR-----VGAGLQNLGNT 119
 Db 121 HPNQSHHNNHHNHQHTSPNLPKPKRVLYPRENIRIGWKQSERKWQVGTGMVNGT 180
 Qy 120 CFANAALQCLTYTTPPLANYMLSHS-HSKTCH-AB--GFCMCTMQAHITQALSPGVDVIK 175
 Db 181 CYLNTSTIQAHLHIIPALANWLVSQAHADCNVABPGSGCIICAMTKILLATQSNQ-SAVR 239
 Qy 176 PMFVINEMRRIARLRFNGQDAHEFLQYTVDMQKACL---NGSNKLDHRHTQATTLVCQ 232
 Db 240 PFLYSLKLOICKHVMVGRQEDAHEFLRFLEAMERAYLMFRNYKELDQLVKETPLGQ 299
 Qy 233 IFGYLSRVKCLNCKGVSTDFPYLDITLBIKAAQSVNKALEOFVPEQLDGENSKCS 292
 Db 300 IFGYLSRVKCLNCKGVSTDFPYLDITLBIKAAQSVNKALEOFVPEQLDGENSKCS 358
 Qy 293 KCKRWVPASKRFTTHRSSNVLTSLKRFANFTGKIAKDVKYPEYLDIRPYMSQ---PNG 349
 Db 359 GCKKKVATQPSLERAPITLCIQKRF-SMIGNKLTQISFKSRIDLSKYAARSQAQA 417
 Qy 350 EPIVYLVYAVLVHTGFGNCHAGHYFCYIKASNGLYQMNDSTVSTDIRSVLSQOAYVLFY 409
 Db 418 QPLTYRLVSMVTHLGASCHGCHYTAIGSTDGSGFYNFDDSVYRPIAMHSVCNTNAYIMFF 477
 Qy 410 -----IRSHDVK-NGBELTHTPSPGQSSRPVLSQRVVTKQAAPGIFGQL 456
 Db 478 ELDSLQAASPAANRPNGVRLTNGHSTTPVPAATVSSPSPT-----RFIGFQL 524
 Qy 457 PSHMIKNPHLN-----GTGPKLDT-----PSSSMSSP 484
 Db 525 PAGGANGYTNQAOKTALQFQKQOQSPONGLOLGTGKFQDTAKPPLVGAHAKGATSA 584
 Qy 485 NGSSVNRASPVNASASVQNVNRSVPIPHPKKQKITTISIHNKLPVRCQSQPNLHNS 544
 Db 585 TANG--NKSSSPSSSSNNHKSINQOQVLPISDDIEDDEMPPRTTAQLPSMPNM--- 639
 Qy 545 SLENTFPVPSSTITNSAVQSTSNASTMSVSKTKPI-----PRSESCSQPMNCKS 597
 Db 640 -TENHTFPAKSPV-----KIQVKTPTVKTPLKSLVPYESASEEAEPLNPRK 686
 Qy 598 KLNSSVLVPYGAESSEDSDESKGLKENGITGVSSHSPQ-----DAED 643
 Db 687 R-----PSGDSSE-SDQES---GQNGHSHKNGSHNGSHNGSHNGSHNGSHNGSH 735
 Qy 644 BEATPHELQPMTLNGANSADSDPKE-----NGLAPDGASCCQOPALHSENPPAK 695
 Db 736 E-----IFKSLKKSADSDDEDDDEEPSIQLTNGWHPPKQS-QSQ-----SK 775
 Qy 696 ANGLPGKLMAPLLSLPEDKILETFR-----LSNK----- 725
 Db 776 APPSP-KTPPSPPAVKSCTGKWKVTRNDEVDAIEDDDVVVVVEGSPVKIPTPNKRNHPF 834
 Qy 726 --LKGSTDEMSPAER 740
 Db 835 SSSKPSSTDSPTPGAKR 851

RESULT 15

Q8VZF5
 ID Q8VZF5 PRELIMINARY; PRT; 661 AA.
 AC Q8VZF5;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE AT3g14400/MLN21.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (Nov-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY064992; AAL5764.1; -;
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 661 AA; 73254 MW; 607DEADDDBC95647 CRC64;

Query Match 18.0%; Score 734; DB 10; Length 661;
 Best Local Similarity 29.7%; Pred. No. 1.1e-43;
 Matches 199; Conservative 114; Mismatches 230; Indels 128; Gaps 21;

Qy 98 KICLKW-----QQTHRVG--AGLQNLGNTCFANAALQCLTYTTPPLANYMLSHSHTC-- 148
 Db 4 KIQSWPSPSSLSQKRNQNGPPLGLNLTCTVNSVLQCLFTTTPPLANFCLTHKSHSHTC 63
 Qy 149 HAEG-----FCMCTMQAHITQALSPGVDVIKPMFVINEMRRIARLRFNGQDAHEFLQY 204
 Db 64 YVDCERKRCDFCIVEKRIARSLSVLTTDAPNKISSCLKIFAEHFKLGRDEAEFLRY 123
 Qy 205 TVDA-----MQKALCNGSNKLDHRHTQATTLVCQIFGYLSRVKCLNCKGVSTDEPY 257
 Db 124 VIDACHNTSLRLKRLRYNGNEPFN---GNSVVKEIFGALQSQVKCLSCGAEKNKADEI 179
 Qy 258 LDITLEIKAAQSVNKALEOFVPEQLDGENSKCSKCKMVPASKRFTTHRSSNVLTSL 317
 Db 180 MDISLEILQSSSVKESLQKFFQSBILDGNNKYRCESCEKLVATARKQMSILQAPNLVQL 239
 Qy 318 KRPFNTGGKIADKVPYLDIRPYMSQNGEP-IVYVLVAVLVHTGFGNCHAGHYFCYI 376
 Db 240 KRFGGIFGGKIDKA-SFGELVLSNFMASKKDPQPEYKLFGLIIVHSGFSPESGHYAYV 299
 Qy 377 KASNGLVQWQNDIVSTDIRSVLSQOAYVLYRSHD-----VKNCGELTHTP 428
 Db 300 KDSLGRWYCCNDSFVLSLSTQEVLSERKAYILFFRSNQRPSAKTLVTSNGLTTSHEVNGC 359
 Qy 429 GQSSPRPVISQVWVTKQAAPGFIGPQLPSHMIKNPHNLNGTGLKDTSPSSMSSPNGNS 488
 Db 360 ETSNPQK-----FIGP-----LNG---FNMKPAEQSFQKGNL 389
 Qy 489 SVNRASPVNASASVQNVNRSVPIPHPKKQKITTISIHNKLPVRCQSQPNLHNSLEN 548
 Db 390 ASSK-----PHK-----FIRPKPRAEQAPLENDLLSSKVEK 420
 Qy 549 -PTKPVPSSTITNSAVQSTSNASTMSVSKTKPIPRSESCSQP--VMNGSKLNSSVLV 605
 Db 421 APLPHAKVSI-----SVNLGAKRVSPVNGRLSFHQDENTAPKANKENSVSVL 468
 Qy 606 PYGAESSEDSDESKGLGKEN-GIGTTIVSSHSPQDAEDEBATPHELQPMTLNGANSAD 664
 Db 469 PTKVNSGTE-----RKFTGTEGNGVKGNGSAPSSNNHKVLAHPHE-----RSNGSS 515
 Qy 665 SDSDPKENGAPDGASCCQOPALHSENPPAKVANGLPKGLMAPLLSLPEDKILETFRLSN 724
 Db 516 NCGDHHKDNLHPGSGNSQNGTAH---PETERNGVSTTQSKGLCSSTKEDPCI-----LLR 568
 Qy 725 KLKGSTDEMSA 735
 Db 569 KDSSRNLEEA 579

Search completed: August 10, 2004, 19:59:49
Job time : 46 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 10, 2004, 19:57:24 ; Search time 21 Seconds
(without alignments)
1905.244 Million cell updates/sec

Title: US-10-049-745-4

Perfect score: 4070

Sequence: 1 MTIVDKASESSDPSAYQNP.....AAESLEPDAASLFFPSEG 775

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:*
- 1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
 - 2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
 - 3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
 - 4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
 - 5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pep.*
 - 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1423.5	35.0	1123	4	US-09-408-865-1
2	1127	27.7	494	3	US-09-019-095A-24
3	889.5	21.9	545	3	US-09-019-095A-38
4	846.5	20.8	521	3	US-09-019-095A-22
5	846.5	20.8	526	3	US-09-019-095A-2
6	841.5	20.7	509	3	US-09-019-095A-9
7	830.5	20.4	508	3	US-09-019-095A-8
8	678	16.7	418	3	US-09-019-095A-10
9	335	8.2	789	1	US-08-338-534-20
10	335	8.2	789	2	US-08-338-534-20
11	335	8.2	789	4	US-09-345-294-20
12	304	7.5	292	4	US-09-406-045-5
13	304	7.5	926	1	US-08-159-340A-2
14	270.5	6.6	2547	3	US-09-058-489-35
15	269.5	6.6	2555	3	US-09-058-489-36
16	247.5	6.1	242	3	US-09-019-095A-23
17	217	5.3	912	1	US-07-789-915A-8
18	217	5.3	912	1	US-08-005-002C-8
19	217	5.3	912	1	US-08-487-203A-8
20	211	5.2	291	4	US-09-167-206-22
21	193.5	4.8	372	4	US-09-406-045-1
22	191	4.7	56	3	US-09-019-095A-15
23	191	4.7	56	3	US-09-019-095A-33
24	186	4.6	462	4	US-09-166-350-18
25	184	4.5	56	3	US-09-019-095A-16
26	172	4.2	907	3	US-08-783-774-2
27	172	4.2	907	4	US-09-328-599A-1

28	172	4.2	907	5	PCT-US95-04611A-19	Sequence 19, Appl
29	170.5	4.2	1291	3	US-09-150-460B-10	Sequence 10, Appl
30	170.5	4.2	1291	3	US-09-220-641-5	Sequence 5, Appl
31	170	4.2	878	4	US-09-556-706B-2	Sequence 2, Appl
32	165.5	4.1	1581	3	US-09-110-517-2	Sequence 2, Appl
33	164	4.0	1939	4	US-09-854-856-48	Sequence 48, Appl
34	164	4.0	1999	4	US-09-854-856-16	Sequence 16, Appl
35	164	4.0	2076	4	US-09-854-856-46	Sequence 46, Appl
36	164	4.0	2136	4	US-09-854-856-14	Sequence 14, Appl
37	162	4.0	1911	4	US-09-854-856-64	Sequence 64, Appl
38	162	4.0	1971	4	US-09-854-856-32	Sequence 32, Appl
39	162	4.0	2048	4	US-09-854-856-62	Sequence 62, Appl
40	162	4.0	2108	4	US-09-854-856-30	Sequence 30, Appl
41	152.5	3.7	1290	3	US-09-150-460B-6	Sequence 6, Appl
42	151.5	3.7	875	3	US-09-150-460B-7	Sequence 7, Appl
43	149.5	3.7	2032	4	US-09-854-856-42	Sequence 42, Appl
44	149.5	3.7	2092	4	US-09-854-856-10	Sequence 10, Appl
45	149.5	3.7	2169	4	US-09-854-856-40	Sequence 40, Appl

ALIGNMENTS

RESULT 1
US-09-408-865-1
; Sequence 1, Application US/09408865A
; Patent No. 6329171
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-51
; CURRENT APPLICATION NUMBER: US/09/408,865A
; CURRENT FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1123
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-408-865-1

Query Match 35.0%; Score 1423.5; DB 4; Length 1123;
Best Local Similarity 37.3%; Pred. No. 1.6e-116;
Matches 347; Conservative 112; Mismatches 238; Indels 233; Gaps 25;

QY	1	MTIVDKASESSDPSAYQNPQSGSEAVSPGMDAGSAGVSSINDV-----	47
Db	1	MPIVDKLEAL-----KPGKDSADDDGELGLASSAKKVLQKIEFPASKSFSYQL	53
QY	48	-----SNHTLSLGPVPGAVVY-SSSVDPKSPQKQDQALGDGIAPPQKVLFPSEKICLK	102
Db	54	EALKSKYVLNPKTEGASRHKSGDDPPARRQSGSEHTVYSCGDGVPAPQKVLFPTELSLR	113
QY	103	WOOTHRVAGLQNLGNTCPANAALQCLTTPPLANTMLSHSEKTKHAEGFCMMCTWQAH	162
Db	114	WERVFRVAGLHNLGNTCFNLATIQCLTTPPLANTMLSHSEKTKHAEGFCMMCTWQAH	173
QY	163	ITQALSNPGDVIKPMFVINEMRRIARHLRFGQEDAEHFLQYTDAMOKACLNGSKLDR	222
Db	174	IYQAFANSNAIKPVSFIRDLKIAIRHFRGQEDAEHFLRYTIDAMQACLNGSKLDR	233
QY	223	HTQATTLVQCIFFGGYLSRVKCLNCKGVSDTDFPYLDITLTKAAQSVNKALFQVKPEQ	282
Db	234	QTQATTLVHQIFGGYLSRVKSVCKSVSDTYDPLDVALEIRQANIVRALELFKADV	293
QY	283	LOGENSYKCKKKMVPASKRTIHRSSNVLTSLKRFANFTGGKIAKDVKYPEYLDIRP	342
Db	294	LSGENAYMACKKXVPASKRFTIHRSSNVLTSLKRFANFTGGKIAKDVKYPEYLDIRP	353
QY	343	YMSQNGEPVIVVLVAVLVHTGNCAGHYFCYIKASGLWYOMNDSIVSTSDIRSVLSQ	402
Db	354	YMSQNGDPVITGLTAVLVHSGYISCHAGHYCYVKASNGQWYOMNDSIVHSSNVKVLNQ	413

Db	61	EWIYTPPLANYMLSRHSQTCRRHKKCMCLTMOAHITRALHNPCHVIQP-----SQAL	111
Qy	187	ARHLRFGNQDEAHEFLQYTVDMOKACLNGSNKLDNRHTQATTLVCOIPGGYLRSVKCLN	246
Db	115	AAGFHRKQDEAHEFLMFTVDAMKCAKLPQHKKQVDHHSKDTTLHIQIPGGYRWSQICLH	174
Qy	247	CKGVSDTFDPYLDITLLEIKAAQSVNKALEQFVPKPEQLDGENSKSCCKKMWPAASKRTFI	306
Db	175	CHGTSDFPYLDIALDIAAQSVOQALEQLVLPKEELGENAYHGCVCLOAPASKTITL	234
Qy	307	HRSSNVLTLSLKRFANFTGGKIAKDVKPYEYLDIRPYMSQNGBPIVVYLYAVLVHTGFN	366
Db	235	HTSAKVILLVLKRFSDVTGNKIAKVQYPECLDMQPYMSQNTGFLVYLYAVLVHAGWS	294
Qy	367	CHAGHYECYIKASNLGYQMNDSTIVSTDTSVLSSQAYVLYPYR-----SHDVXNGG	419
Db	295	CHNGHYFSYKVAQEGQWYXMDDAEVTASSITSVLSQAYVLYFYTKSEWERHSESVSRGR	354
Qy	420	ELHTPTSPGCGSSRPVLSQVVTNKAQAPGI-----GPQLPSHMKNPPLHNGTG	471
Db	355	E-----PVALGAED--TDRRTQGBELKXDHFCLOAPELDEHLVERATQESTLD	400
Qy	472	PLKDTPPSSSMSSPNAN--SSVNRASPVNASASVQNVMSVNRSSVYPEHPKKQKTTISIHNKL	530
Db	401	HWKFLQEQNKTKPEFNVRKVEGTLP--PDVLVLIHQSKYKCGMKNHHPQEQSSLLLSSTT	458
Qy	531	PVRQ	534
Db	459	PTHQ	462

RESULT 3
US-09-019-095A-38
. Sequence 38. Application US/09019095A

```

; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCl-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1995-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 38
; LENGTH: 545
; TYPE: PRT
; ORGANISM: murine
; US-09-019-095A-38

```

[illegible]

Db	177	LCQGTSDTYDRFLDIPDISSAQSQYQALWDTKSEELCGDNAYCYGCKRQKWPASKTLIH	236
Qy	306	IHRSSNVLTLSLKEFANFTGKIAKDVKYPEYLDIRPYMSQNGEPIVTVLVAVLVHTGF	365
Db	237	VHIAPKVLWVVLNRFSAFTGNKLDKQSYPEFLDLKPYLSEPTGGPLPALYAVLVHDGA	296
Qy	366	NCHAGHYFCYIKASGLWYQWNSIVSTDTRSVLSQOAYVLYFR	418
Db	297	TSHSGHYFCCVKAGHGKWKYKDDTKVTRCDVTSVLNENAYLVFYQQANLKOVSDMPEG	356
Qy	419	--GELTHPTH-----SPGOSSPRPVTSQ-----RVVTNKO	446
Db	357	RINEVLDPYQLKKSRRKKKKKSPFTEDLGEPCENRDKRAIKETSLGKGVLQVNHKK	416
Qy	447	AAPGF----TGPQLPSHWIKNPPHLNGTGPLKDTFSSS--MSSPMNGNSVNRASP----	495
Db	417	AGQKHGNTKMPQKQNHQ--KAGQNLNRNTEVELDLDPADAIVIHQPRSTANWGRDSDPKENQ	475
Qy	496	-----VNASASVQNVSNR	509
Db	476	PLHNADRLTTSOGPVNTWOLCR	497

RESULT 5

```

US-09-019-095A-2
; Sequence 2, Application US/09019095A
; Patent NO. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 526
; TYPE: prt
; ORGANISM: Murine
US-09-019-095A-2

```

Query Match 20.8%; Score 846.5; DB 3; Length 526;
Best Local Similarity 37.1%; Pred. No. 4.8e-66;
Matches 186; Conservative 89; Mismatches 153; Indels 75; Gaps 12

Db	8	PEADPALSPDAPELHQDEAAVVVEELTVNGKHSLSWESPQGGCLQNTGNSCYLNAALQ	67
Qy	128	CITYTPPLANYMLSHESKTKCHAEAGFCNMCTMOAHITQAL--SNPGDVIKPFMFINEMRR	185

68	CLTHPTPLADYMLISQEHSSOTCCSPGEGCKLCAMEALVTQSLHSHSGDWKPFSHIITS---	124
186	IARHLRPGNOQDAHEFFLOYVDAMQACLINGSNKLDRHTQATTLVCOIFGGYILRSVKCL	245
125	-AFPHK--QOQDAHEFLMFTLETMHESCLOVHRQSKPTSEDSPIHIDIFGGWNRSQIKL	181
246	NCKGVSDFTRPDIJLTLEIKAAQSVNKALQFVKPEQLDGENSYKCSCKKQWVPASKFT	305
182	LCQGTSDTYDRFLDIPJLIDISAAQSVQKALMDTEKSEBLCGDNAYYCGKRCROKPFASKLH	241
306	IHRSSNVLTLSLRFANFTCGKIADKYKYPEYLDIRPYMOPNGEPIVYLYAVILVHTGF	365
242	VHTAPKVLMLVNLRFSAFTGNKLDKRVSYPEFLDKPYLSEPTGGPUPYALYAVILVHDGA	301

```

366 NCRAGHYFCYKASGLWYQMNDSIVSTSDIRSVLSQOAYVLFYLR-----SHDVXNG 418
302 TSHSGHYFCCKAGHGKWKYKMDTKVTRCDVTSVLINAYVLFYQOANLKQVSDMPEG 361
419 --GELTHPTH-----SPGSSSRPVISQ-----RVVTNKQ 446
362 RINEVLDPEYQLKKRRKKKKKSPFTEDLGEPCENRDKRAIKETSLGKGKVLQEVNHKK 421
447 AAPGF---IGPOLPSHMIKNPHHLNGTGLPKDTPSSS--MSSPNGNSGVNRASP-----495
422 AGQKHGNTKMLPQKNHC-KAGQNLNRTVEVLDLPADAIVHQPRSTANWGRDSDKENQ 480
496 -----VNASASVQNVSNVR 509
481 PLHNADRLLTSGQPVNTWQLCR 502

RESULT 6
US-09-019-095A-9
; Sequence 9, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; APPLICANT: Zhu, Yuan
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; TITLE OF INVENTION: Cell Growth
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019,095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Murine
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(509)
; NAME/KEY: VARIANT
; LOCATION: (1)...(509)
; OTHER INFORMATION: Xaa = Any Amino Acid
; US-09-019-095A-9

```

US-09-019-095A-10
; Sequence 10, Application US/09019095A
; Patent No. 6287858
; GENERAL INFORMATION:
; APPLICANT: D'Andrea, Alan D.
; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
; FILE REFERENCE: DFCI-435p2A2
; CURRENT APPLICATION NUMBER: US/09/019, 095A
; CURRENT FILING DATE: 1998-02-05
; PRIOR APPLICATION NUMBER: PCT/US96/12884
; PRIOR FILING DATE: 1996-08-07
; PRIOR APPLICATION NUMBER: US 60/002,066
; PRIOR FILING DATE: 1995-08-09
; PRIOR APPLICATION NUMBER: US 60/019,787
; PRIOR FILING DATE: 1996-06-14
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 418
; TYPE: PRI
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: consensus
US-09-019-095A-10

Query Match 16.7%; Score 678; DB 3; Length 418;
Best Local Similarity 38.3%; Pred. No. 2.6e-51;
Matches 158; Conservative 73; Mismatches 114; Indels 68; Gaps 18;
QY 101 LKWOQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYMLSHEHSKTCCHAEFGCMCTQW 160
DB 26 LSW-ESPFGCGLQNTGNSCYNALQCLTHPPADYMLSQHSOTCCSPGC-KCAME 83
QY 161 AHITQALSNPGDVIKPMFVINEMRRIARHLRFGNQEDAHEFLQYTVDMQKACLNGSNKL 220
DB 84 --VTQSL--GDVWKPISLTSAFHK-----HQEDAHEFLMFLTETHESCL--QV 128
QY 221 DRHTQATLVCOIFGGYLSRVKCLNCKGVSDTFDPYDLITLIEKAAQSVNKALQFVKP 280
DB 129 HRQSPDTPSIHDFGGMWRSAQKCLGTS--TDFLDP-----LDISSAQSVQALWD--TKS 179
QY 281 EOLDGENSKCKKQKVPASRFTIHRSSNVLTLSLKRFANFTGKIAKQVYKPYLDI 340
DB 180 EELG--NAYCG-CRQKMPASTLHVIAPKVLVL--RPSAFTGNKLDKRVSYPEFLDL 233
QY 341 RPYMSQNGEPIVYVLYAVLVHTGFNCHAGHYFCYIKASNGLWYQMNDSIVSTSDIRSVL 400
DB 234 KPYLSEPTGGPLPYALYAVLVHDGAT--SSGHYFCCVKAGHGWKYMDDTKVTRCDVTSVL 292
QY 401 SQQAVLVFIR-----SHDVKNKGELTHP-----THSPGQSSPR 434
DB 293 NENAYLVFYQQAQKQVSDIMPEGREVLDPYQLKSRKHKHGEDRLGEVNHKAGHNTK 352
QY 435 PVISORVVTNKQAAPGF---IGPOLPSHMIKNPPLHNGTGPLKOTPSSMS 483
DB 353 --LPQO---NHQAGNLRNTEVELDPLDIVIHOPRSTANWNGSKENQPNRLTS 400

RESULT 9
US-08-431-080-20
; Sequence 20, Application US/08431080
; Patent No. 5698686
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston

STATE: TEXAS
COUNTRY: UNITED STATES OF AMERICA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/431,080
FILING DATE: Concurrently Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SN 08/326,781
FILING DATE: October 20, 1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:155/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (713) 789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 789 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-431-080-20

Query Match 8.2%; Score 335; DB 1; Length 789;
Best Local Similarity 23.4%; Pred. No. 1.6e-20;
Matches 112; Conservative 74; Mismatches 173; Indels 120; Gaps 16;
QY 96 SEKICLKW--QOTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYML-----SHEHSKTC 148
DB 330 STRIVKNWGDKFTNLKPRGLNHGVTCTYNAAVQAMLHIPSIOHYLFDILMCKYDSTISK 389
QY 149 HAEGFCMCTQAHITQALSNP-----GDVIKPMFVINEMRRIARHLRFGNQEDAHEFLQ 203
DB 390 NSVSYTLAETSKMMLPVSKNPKVSAIYNPKHLISRLDDINCMMSEWQEDSHEYFM 449
QY 204 YTVDMQKACLNGSNKLDRHTQATLVCOIFGGYLSRVKCLNCKGVSDTFDPYDLITLIE 263
DB 450 SLMSRLQEDSVKPGHKL-----IESIIVDFGGLLKQIVTCKSCGISKTEQPFYDLSLH 504
QY 264 IKAAG----- 268
DB 505 LKGGKKLDPNSDLSDSINGTSATTTSTTSNAATKPSLSSSSSVNLNNGSPFAAASDLSS 564
QY 269 -----SVNKALQFQVKPE--QLDGE--NSYKCSKCKKQKVPASRFTIHRSSNVLTLSLKRF 320
DB 565 ANRRSIEKSIKDFNPPELLIKVDKEQKGVCEKCHKTNAVKHSSILRAPETLLVHLKPF 624
QY 321 ANFTG---GKIAQVYKPYEYLDIRPYMSQNGEPIVYVLYAVLVHTGFNCHAGHYFCYIK 377
DB 625 -RFNGTSSSKMKQAVSYPMFLDLTEY-CESKELPKVYQLLSVVVHEGRSLSSGHYIAHCK 682
QY 378 ASNGLWYQMNDSIVSTSDIRSVLSQ-QAYVLPIRSHDVKNKGELTHPTSPGQSSPRPV 436
DB 683 QPDGGSWATYDDEYINIIISERDVLKEPNAYLLYTR-----LT-----PKSVPLPL 727
QY 437 ISORVVTNKQAAPGFIGPOLPSHMIKNPPLHNGTGPLKOTPSSM-----SSPN 485
DB 728 AKSAAAT-----GNVTSKSKQEQAVNEPN---NRPLKINSKKNRKKWKIKKKGSSPN 777
RESULT 10
US-08-938-534-20
; Sequence 20, Application US/08938534
; Patent No. 5916752

```

; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,534
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-938-534-20

```

```

Query Match      8.2%; Score 335; DB 2; Length 789;
Best Local Similarity 23.4%; Pred. No. 1.6e-20;
Matches 112; Conservative 74; Mismatches 173; Indels 120; Gaps 16;

Qy 96 SEKICLKW--QOTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYML-----SHEHSKTC 148
Db 330 STRIVKNGDKFTNLKPRGLNHGVTCTNAAVQAMLHPSIQHYLFIDILMGKYDSTISK 389

Qy 149 HAEGFCMMCTQAHITQALSNP-----GDVIKPMFVINEMERIAHRLFNGQDAHEFLQ 203
Db 390 NSVSYTLAETSCKMWLPVSKNPKNVASYNPKHLISRLDDINCMSEWOQDSHEYFM 449

Qy 204 YTDAMQKACLNGSKLDRHTQATTLVCQIFGGYLSRVKCLNCKGVSDTDPDYLIDPLE 263
Db 450 SLMSRLQEDSVPKGHKL-----IESIYDIIFGGLLKQIVCTCKSCGISKTEQPFYDLSLH 504

Qy 264 IKAQAQ----- 268
Db 505 LKGGKKLDPNGLSDSDSINGTSATTTSTTNAATKPSLSLSSSSSVNLNNGSPFAAASDLSS 564

Qy 269 -----SVNKALEQFVKPE--QLDGE-NSYKCSCKCKMYPASKRFTIHRSSNVLTLSLKRF 320
Db 565 ANRFSIEKSKIDFPNPELIIKVDKEQKGVCEKCHKTNAVKHSSILRAPETLLVHLKRF 624

Qy 321 ANFTG---GKIAKDVKYPEYLDIREYMSQNPGEPIVYVLYAVLVHTGFNCCHAGHYFCYIK 377
Db 675 -RFNGTSSSKMKQAVSPMFLDLTEY-CESKELPVKQYLLSVVWHEGRSLSSGHYIAHCK 682

Qy 378 ASNGLWYQNDISVSTSDIRSVLSQ-QAYVLFIYRSHDVKNKGELTHPTHPGQSSPRPV 436
Db 683 QPDGSWATYDDYINIISERDVLKEPNAYLLYTR-----LT-----PKSVLPLEL 727

```

```

Qy 437 ISQVVVTKQAAPCFIGPOLPSHMKNPPLHNGTGPLKDTPESSM-----SSPN 485
Db 728 AKSAMAT-----GNVTSKSKQEQAVNEPN--NRPLKINSKKNRKKKKKKSSPN 777

RESULT 11
US-09-345-294-20
; Sequence 20, Application US/09345294
; Patent No. 6387619
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/345,294
; FILING DATE: 30-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 789 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:

US-09-345-294-20

```

```

Query Match      8.2%; Score 335; DB 4; Length 789;
Best Local Similarity 23.4%; Pred. No. 1.6e-20;
Matches 112; Conservative 74; Mismatches 173; Indels 120; Gaps 16;

Qy 96 SEKICLKW--QOTHRVGAGLQNLGNTCFANAALQCLTYTPPLANYML-----SHEHSKTC 148
Db 330 STRIVKNGDKFTNLKPRGLNHGVTCTNAAVQAMLHPSIQHYLFIDILMGKYDSTISK 389

Qy 149 HAEGFCMMCTQAHITQALSNP-----GDVIKPMFVINEMERIAHRLFNGQDAHEFLQ 203
Db 390 NSVSYTLAETSCKMWLPVSKNPKNVASYNPKHLISRLDDINCMSEWOQDSHEYFM 449

Qy 204 YTDAMQKACLNGSKLDRHTQATTLVCQIFGGYLSRVKCLNCKGVSDTDPDYLIDPLE 263
Db 450 SLMSRLQEDSVPKGHKL-----IESIYDIIFGGLLKQIVCTCKSCGISKTEQPFYDLSLH 504

Qy 264 IKAQAQ----- 268
Db 505 LKGGKKLDPNGLSDSDSINGTSATTTSTTNAATKPSLSLSSSSSVNLNNGSPFAAASDLSS 564

Qy 269 -----SVNKALEQFVKPE--QLDGE-NSYKCSCKCKMYPASKRFTIHRSSNVLTLSLKRF 320

```


Db 565 ANRRPSIEKSIDPENPELIKVDKEQGYVCEKCHKTNNVAKVHSSILRAPETLLVHLKKF 624
Qy 321 ANFTG--GKIADVKYPEYLDIRYMSQPNGEPIVYVLYAVLVHTGPNCHAGHYFCVIK 377
Db 625 -RFGNTSSKMQAVSYPFMDLTRY-CESKELPVKYQLLSVVVHEGRSLSGHYIAHCK 682
Qy 378 ASNGIWMQNDISVSTSDIRSVLSQ-QAYVLFYIRSHVDKNGGELTHPSPGQSSRPV 436
Db 683 QPDGSMATYDEYINISERDVLKPNAYLLYTR-----LT-----PKSVPLPL 727
Qy 437 ISQRVTVKQAAPGFIGQLPSPHMKPNPHLNGTGLKMDTPSSM-----SSPN 485
Db 728 AKSAMAT-----GNVTSKSQEQAVNEPN--NRPKINSKKNRKKWKIKKGSSN 777

RESULT 12
US-09-406-045-5
; Sequence 5, Application US/09406045
; Patent No. 6451994
; GENERAL INFORMATION:
; APPLICANT: Kapeller-Libermann, Rosana
; TITLE OF INVENTION: 23413, A NOVEL HUMAN UBIQUITIN
; FILE REFERENCE: 5800-50
; CURRENT APPLICATION NUMBER: US/09/406,045
; CURRENT FILING DATE: 1999-09-27
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: ProDom consensus sequence
US-09-406-045-5

Query Match 7.5%; Score 304; DB 4; Length 292;
Best Local Similarity 30.2%; Pred. No. 1.7e-18;
Matches 88; Conservative 52; Mismatches 103; Indels 48; Gaps 13;
Qy 103 WQOTHRVG-AGLQNLGNTCFANAALQCLTYTPPLANYMLSHEH-SKTHAEGFCM----- 155
Db 2 WDSKRGPGVTGLKNLGNTCYMNSVLQCLYHVPLPREYFLEDEYSEVMNNSNPLGMKE 61
Qy 156 MCTMQAHT-QALSNGPD-VIKPMFVINEMRIARHLRFG-NOEADHPELOYTVD----- 207
Db 62 LATAYAKLVHQMWSNSKNSVAPTOFLTITVGKFSQFSGYQQQDSQBFKLQDDAHED 121
Qy 208 --AMOKACLAGNSNKLDRHTOATTLV-----CQIFGGYLRSRVKCLN 246
Db 122 FNSLMKPYVEBOVK-DSNEKSTALVNVSEAEWENHKKRNDISIITDIFQGFKSIKCP 180
Qy 247 CKGVSD-TFDPYLDITLRIKA-----AQSVNKALQFVKPEQLDGENSYKSKCKMWP 300
Db 181 CEHTSETTFEPMDLSLPIPSADNHQNLQCLSEFTKKTLEGDNKWKYCPKCKKQEA 240
Qy 301 SKRFTIHRSSNVLTLSLKRFPAN-----FTGCKTAKDVKYP-EYLDIRPYMSQ 346
Db 241 TKKLDIWLKPPVLVILHKKRFSYDROWGRDRDKLTTVEFPLEDLDMSPYVDK 291

RESULT 13
US-08-159-340A-2
; Sequence 2, Application US/08159340A
; Patent No. 5565352
; GENERAL INFORMATION:
; APPLICANT: Hochstrasser, Mark
; APPLICANT: Papa, Feroz
; TITLE OF INVENTION: DEUBIQUITINATING ENZYME: COMPOSITIONS
; TITLE OF INVENTION: AND METHODS
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433

; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/159,340A
; FILING DATE: 24-NOV-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: ARCD:112/HYL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 926 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-159-340A-2

Query Match 7.5%; Score 304; DB 1; Length 926;
Best Local Similarity 26.6%; Pred. No. 1.1e-17;
Matches 102; Conservative 57; Mismatches 126; Indels 98; Gaps 18;
Qy 112 GLQNLGNTCFANAALQCLTYTPPLANYMLSHEHSTKTH-----AEGFC-----MM 156
Db 563 GLENLGNSCYMNCIIQCLILGTHLTIQIFLDDSYAKHININSKLGSKGILAKYFARLVHMM 622
Qy 157 CTMQAHTIQAALSNPGDVIKPMFV-----INEMRIARHLRFGNOEADHPELOYTVDAMQ 210
Db 623 YKEQVDGSKKIS-----ISPIKFKIACGSVNSLFKTA-----SQDCCQEFQCLLDGLH 671
Qy 211 KACLN--GSN-----KLDHRHTOAT-----TLVCQIFGGYLRSRV 242
Db 672 ED-LNQCCSNPPLKLSQEAERREKLSIRTASSIEWERFLTTFDSVIVDLFQGYASRL 730
Qy 243 KCLNCKGVSDTFDPYLDITLRIKAQSVNK-----ALQFVKPEQLDGENSYKSKCKKM 297
Db 731 KCKVCSHTSTTYQPTVLSIPIPKNSRNNTIEDCFREFTKCNLEVDQWMLCPHCEKR 790
Qy 298 VPASKRFTIHRSSNVLTLSLKRFPANFTGCKTAKDVKYPEYLDIRPYMSQ-----PNG-- 349
Db 791 QBSTKQLTITLPRNLIVHLKRFNLL-NKNNDFFVIYFPLDLPFWANDFGVFPFGVN 849
Qy 350 -----EPVVVLYAVLVHTGFCNCHAGHYFCYIKASNGL-----WYQMNDSD-- 392
Db 850 DDELPQIRQIPFKYELXGVACHFG-TLYGGHYTAYK--KGLKGLWLYFDDTKYKPVKN 906
Qy 393 TSDIRSVLSQAYVLFYIRSHDV 415
Db 907 KAD--AINSNAVLYFHYRVGV 926

RESULT 14
US-09-058-489-35
; Sequence 35, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; TITLE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WHI97-08pA

RESULT 1

Q802X0	PRELIMINARY;	PRT;	1016 AA.
ID	Q802X0		
AC	Q802X0;		
DT	01-JUN-2003 (TrEMBLrel. 24, Created)		
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Similar to KIAA1453 protein (Fragment).		
OS	Brachydanio rerio (zebrafish) (Danio rerio).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;		
OC	Cyprinidae; Danio.		
OX	NCBI_TaxID=7955;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-AB;		
RA	Strausberg R.;		
RL	Submitted (Feb-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC047168; RAH47168.1; --		
DR	GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.		
DR	GO; GO:0004379; F:glycylpeptide N-tetradecanoyltransferase ac. . .; IEA		
DR	GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.		
DR	GO; GO:0006499; P:N-terminal protein myristoylation; IEA.		
DR	GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.		
DR	InterPro; IPR003006; IG_MHC.		
DR	InterPro; IPR000903; NMT.		
DR	InterPro; IPR001394; Peptidase_C19.		
DR	Pfam; PF00443; UCH; 1.		
DR	PROSITE; PS00290; IG_MHC; 1.		
DR	PROSITE; PS00976; NMT_2; 1.		
DR	PROSITE; PS00972; UCH_2_1; 1.		
DR	PROSITE; PS00973; UCH_2_2; 1.		
DR	PROSITE; PS02335; UCH_2_3; 1.		
FT	NON TER 1016 1016		
SEQ	SEQUENCE 1016 AA; 113138 MW; 05AFD86206CC9DAA CRC64;		
Query Match 2.2%; Score 17; DB 13; Length 1016;			
Best Local Similarity 100.0%; Pred. No. 6.8e-08;			
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps			
Qy	232 QIFGGYLRSVKCLNCK 248		
Db	203 QIFGGYLRSVKCLNCK 219		
RESULT 2			
Q9H9C5	PRELIMINARY;	PRT;	548 AA.
ID	Q9H9C5		
AC	Q9H9C5;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
DE	Hypothetical protein FLJ12851.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,		
RA	Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,		
RA	Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,		
RA	Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,		
RA	Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.;		
RT	"NEDO human cDNA sequencing project.;"		
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK022913; BAB14306.1; --		
DR	GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.		
DR	GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.		
DR	GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.		
DR	InterPro; IPR003006; IG_MHC.		
DR	InterPro; IPR001394; Peptidase_C19.		
DR	Pfam; PF00443; UCH; 1.		
DR	PROSITE; PS00290; IG_MHC; 1.		

DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 548 AA; 60332 MW; 2C5ED382105AC2C6 CRC64;
 Query Match 1.8%; Score 14; DB 4; Length 548;
 Best Local Similarity 100.0%; Pred. No. 4.8e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 310 SNVLTLSLKRFAF 323
 Db 321 SNVLTLSLKRFAF 334
 RESULT 3
 Q8BV95 ID Q8BV95 PRELIMINARY; PRT; 180 AA.
 AC Q8BV95;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE CDNA FLJ12851 FIS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The PANTOM Consortium,
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK079342; BAC37614.1; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH_1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 SQ SEQUENCE 180 AA; 19513 MW; BC65A8C66379DEEA CRC64;
 Query Match 1.5%; Score 12; DB 11; Length 180;
 Best Local Similarity 100.0%; Pred. No. 0.002;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 QCLTYTPPLANY 138
 Db 138 QCLTYTPPLANY 149
 RESULT 4
 Q8IXW9 ID Q8IXW9 PRELIMINARY; PRT; 285 AA.
 AC Q8IXW9;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to KIAA1453 protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Blood;
 RA Strausberg R.;
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC038983; AA38983.1; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH_1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 285 AA; 31791 MW; 23AA44DD2217342E CRC64;
 Query Match 1.5%; Score 12; DB 4; Length 285;
 Best Local Similarity 100.0%; Pred. No. 0.0031;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 127 QCLTYTPPLANY 138
 Db 138 QCLTYTPPLANY 149
 RESULT 5
 Q9FJL7 ID Q9FJL7 PRELIMINARY; PRT; 829 AA.
 AC Q9FJL7;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similarity to deubiquitinating enzyme.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kotani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 Sequence features of the regions of 1,367,185 bp covered by 19
 Physically assigned P1 and TAC clones."
 RL DNA Res. 5:203-216(1998).
 DR EMBL; AB013396; BAB08869.1; -.
 DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH_1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 829 AA; 91549 MW; 71CBA02EBCAED055 CRC64;
 Query Match 1.5%; Score 12; DB 10; Length 829;
 Best Local Similarity 100.0%; Pred. No. 0.0082;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 110 GAGLQNLGNTCF 121
 Db 106 GAGLQNLGNTCF 117
 RESULT 6
 Q9FPS4 ID Q9FPS4 PRELIMINARY; PRT; 859 AA.
 AC Q9FPS4;
 DT 01-MAR-2001 (TrEMBLrel. 15, Created)
 DT 01-MAR-2001 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 23, Last annotation update)
 DE Ubiquitin-specific protease 23.
 GN UBP23
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

```

X NCBI_TaxID=3702;
N [1]
P SEQUENCE FROM N.A.
MEDLINE=20567829; PubMed=11115897;
A Van N., Doelling J.H., Falbel T.G., Durski A.M., Vierstra R.D.;
T "The Ubiquitin-Specific Protease Family from Arabidopsis. AtUBP1 and 2
L Are Required for the Resistance to the Amino Acid Analog Canavanine.";
R Plant Physiol. 124:1828-1843(2000).
R ENBL; AF302671; AA424761.1; -.
R GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
R GO; GO:0008233; F:peptidase activity; IEA.
R GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
R GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
R InterPro: IPR001394; Peptidase_C19.
R Pfam: PF00443; UCH_1.
R PROSITE; PS00972; UCH_2_1; 1.
R PROSITE; PS00973; UCH_2_2; 1.
R PROSITE; PS02035; UCH_2_3; 1.
R Protease.
R SEQUENCE 859 AA; 94889 MW; 4CC8E082B54C4FE9 CRC64;
Q Query Match 1.5%; Score 12; DB 10; Length 859;
Best Local Similarity 100.0%; Pred.No. 0.0084;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GAGLQNLGNTCF 121
DB |||||
106 GAGLQNLGNTCF 117

RESULT 7
Q84P80 PRELIMINARY; PRT; 842 AA.
AC
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DT Putative ubiquitin-specific protease (Fragment).
DE Oryza sativa (japonica cultivar-group).
OS Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=30843;
RX MEDLINE=22584437; PubMed=12684538;
RA Cooper B., Clarke J., Budworth P., Kreps J., Hutchison D., Park S.,
RA Guimil S., Dunn P., Luginbuhl P., Ellero C., Goff S.A., Glazebrook J.;
RT "A network of rice genes associated with stress response and seed
RT development.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:4945-4950(2003).
RL ENBL; AY224488; AA072607.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro: IPR001394; Peptidase_C19.
DR Pfam: PF00443; UCH; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS02035; UCH_2_3; 1.
DR Protease.
DR NON_TER 1
FT SEQUENCE 842 AA; 93153 MW; F3BFCTAAD5E93620 CRC64;
SQ

Query Match 1.4%; Score 11; DB 10; Length 842;
Best Local Similarity 100.0%; Pred.No. 0.089;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 GAGLQNLGNTC 120
DB |||||
1 GAGLQNLGNTC 11

RESULT 9
Q9MAQ3 PRELIMINARY; PRT; 887 AA.
AC
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CDS.
DE F9111.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.P.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Bei, Buehler E., Chin C.,
RA Chlou J., Choi E., Dunn P., Gonzalez A., Howng B., Kim C., Koo T.,
RA Lee J.M., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Pham P., Sakano H., Schwartz J., Shinn P., Thaveri A., Toriumi M.,
RA Vaysberg M., Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL ENBL; AC006424; AAF31287.1; -.
RL PIR; C86453; C86453.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

```

```

RESULT 8
Q9HGT8 PRELIMINARY; PRT; 132 AA.
AC
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DT Ubiquitin carboxyl-terminal hydrolase (Fragment).
DE Coccidioides immitis.
OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Onygenales; mitosporic Onygenales; Coccidioides.
OX NCBI_TaxID=5501;
RN [1]
RP SEQUENCE FROM N.A.
RA Delgado N., Cole G.T.;
RT "Identification of differentially expressed genes in Coccidioides
RT immitis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF288062; AAG00799.1; -.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0004221; F:ubiquitin thiolesterase activity; IEA.
DR GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
DR InterPro: IPR001394; Peptidase_C19.
DR InterPro: IPR000626; Ubiquitin.
DR Pfam: PF00240; ubiquitin; 1.
DR Pfam: PF00443; UCH; 1.
DR SMART; SM00213; UBQ; 1.
DR PROSITE; PS00053; UBIQUITIN_2; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR Hydrolase.
DR NON_TER 1
FT SEQUENCE 132 AA; 14297 MW; 151651504DE67968 CRC64;
SQ

Query Match 1.3%; Score 10; DB 3; Length 132;
Best Local Similarity 100.0%; Pred.No. 0.18;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 AGLQNLGNTC 120
DB |||||
108 AGLQNLGNTC 117

RESULT 9
Q9MAQ3 PRELIMINARY; PRT; 887 AA.
AC
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CDS.
DE F9111.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.P.,
RA Altafi H., Nguyen M., Lam B., Southwick A., Bei, Buehler E., Chin C.,
RA Chlou J., Choi E., Dunn P., Gonzalez A., Howng B., Kim C., Koo T.,
RA Lee J.M., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,
RA Pham P., Sakano H., Schwartz J., Shinn P., Thaveri A., Toriumi M.,
RA Vaysberg M., Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RA Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RL ENBL; AC006424; AAF31287.1; -.
RL PIR; C86453; C86453.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.

```

DR GO:0004221; Fubiquitin thiolesterase activity; IEA.
 DR GO:0006512; P:ubiquitin cycle; IEA.
 DR GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR006615; DUSP.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR SMART; SM00695; DUSP; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 SQ SEQUENCE 887 AA; 100619 MW; 419F8E5E067FA7B2 CRC64;

Query Match 1.3%; Score 10; DB 10; Length 887;
 Best Local Similarity 100.0%; Pred.No.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLQNLGNTCF 121
 |||||
 Db 297 GLQNLGNTCF 306

RESULT 10

Q9C585
 ID Q9C585 PRELIMINARY; PRT; 901 AA.
 AC Q9C585;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Ubiquitin-specific protease-like protein.
 GN AT5G22030.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL589883; CAC34496.1; -
 DR GO:0004197; F:cysteine-type endopeptidase activity; IEA.
 DR GO:0006233; F:peptidase activity; IEA.
 DR GO:0004221; F:ubiquitin thiolesterase activity; IEA.
 DR GO:0006511; P:ubiquitin-dependent protein catabolism; IEA.
 DR InterPro; IPR001394; Peptidase_C19.
 DR Pfam; PF00443; UCH; 1.
 DR PROSITE; PS00972; UCH_2_1; 1.
 DR PROSITE; PS00973; UCH_2_2; 1.
 DR PROSITE; PS0235; UCH_2_3; 1.
 KW Protease.
 SQ SEQUENCE 901 AA; 101704 MW; 5778D09590C87886 CRC64;

Query Match 1.3%; Score 10; DB 10; Length 901;
 Best Local Similarity 100.0%; Pred.No.1;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 GLQNLGNTCF 121
 |||||
 Db 310 GLQNLGNTCF 319

Search completed: August 10, 2004, 20:16:37
 Job time : 44 secs

Pred. No. is the number of results predicted by chance to have a

QY	1981	CCGGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCCATGACC	2040
Db	1981	CCCGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCCATGACC	2040
QY	2041	CTAAACGGTGCTTAATAGTGCACAGACGACAGTCAACCGAAGAAAAACGGCGCTTAGCGCCT	2100
Db	2041	CTAAACGGTGCTTAATAGTGCACAGACGACAGTCAACCGAAGAAAAACGGCGCTTAGCGCCT	2100
QY	2101	GATGTTGCCAGCTGCCAAGGCCAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCTTAAGGCA	2160
Db	2101	GATGTTGCCAGCTGCCAAGGCCAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCTTAAGGCA	2160
QY	2161	AACGCTCTTCTCTGGAAGCTTGATGCTGCTCTTCTGCTGCTCTCCAGAAACAAAAATC	2220
Db	2161	AACGCTCTTCTCTGGAAGCTTGATGCTGCTCTCTTCTGCTGCTCTCCAGAAACAAAAATC	2220
QY	2221	TTAGAGACCTTCAGGCTTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGACCT	2280
Db	2221	TTAGAGACCTTCAGGCTTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGACCT	2280
QY	2281	GGAGCAGAGAGGGGCCCTCCCGAGGACCGGACGCCGAGCTCAGCTCGACGCCCCGCC	2340
Db	2281	GGAGCAGAGAGGGGCCCTCCCGAGGACCGGACGCCGAGCTCAGCTCGACGCCCCGCC	2340
QY	2341	GCCGAATCCCTCGAGGACCCAGATGCGGCGCGCAAGCTTATTCCTTTAGTAGGGTTAA	2400
Db	2341	GCCGAATCCCTCGAGGACCCAGATGCGGCGCGCAAGCTTATTCCTTTAGTAGGGTTAA	2400
QY	2401	TTTTAGCTTGCA	2412
Db	2401	TTTTAGCTTGCA	2412
RESULT 2			
LOCUS	AX786887	4659 bp	DNA linear PAT 17-JUL-2003
DEFINITION	Sequence 22 from Patent WO0238744.		
ACCESSION	AX786887		
VERSION	AX786887.1	GI:32954188	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Lee, E.A., Hafalia, A.J., Yue, H., Lal, P.G., Yao, M.G., Lu, Y., Chawla, N.K., Warren, B.A., Lu, D.A., Baughn, M.R., Delegeane, A.M., Burford, N., Borowsky, M.L., Lee, S., Xu, Y., Griffin, J.A., Kallick, D.A., Gandhi, A.R., Arvizu, C.S., Ison, C.H., Tang, Y.T., Azimail, Y., Elliott, V.S., Swarnakar, A., Ramkumar, J., Nguyen, D.B., Tribouley, C.M., Lo, T.P., Au-Young, J., Thangavelu, K. and Kearney, L.		
TITLE	Proteases		
JOURNAL	Patent: WO 0238744-A 22 16-MAY-2002;		
FEATURES	Incyte Genomics, Inc. (US)		
source	Location/Qualifiers		
	1. .4659		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	/note="Incyte ID No: 7997881CB1"		
ORIGIN			
Query Match	98.3%;	Score 2371.6;	DB 6; Length 4659;
Best Local Similarity	99.8%;	Pred. No. 0;	
Matches 2374; Conservative	0;	Mismatches 4;	Indels 0; Gaps 0;
QY	1	GGGCTGTGCGCGCGCGCGCGCGCGCGCGCGGATGGAGCGCGCGCGCGCGCGGT	60
Db	83	GGGCTGTGTCGCGCGCGCGCGCGCGCGCGCGCGGATGGAGCGCGCGCGCGGT	142
QY	61	CAGAGTTGAACATGACCATAGTTCGACAAAGCTTCTGATCTTCAGACCCATAGCCTAT	120
Db	143	CAGAGTTGAACATGACCATAGTTCGACAAAGCTTCTGATCTTCAGACCCATAGCCTAT	202

QY	1201	AAAGCTAGCAATGGCCCTCTGGGTATCAAAATGAATGACTCCATTTGTATCTACAGTGATATT	1260
DB	1283	AAAGCTTAGCNAATGGCCCTCTGGGTATCAAAATGAATGACTCCATTTGTATCTACAGTGATATT	1342
QY	1261	AGATCGGTACTCAGGCCAAACAAGCCTAATGTGCTCTTTTATATCAGTCCGATGATGTGAAA	1320
DB	1343	AGATCGGTACTCAGGCCAAACAAGCCTAATGTGCTCTTTTATATCAGTCCGATGATGTGAAA	1402
QY	1321	AATGGAGGTGAACCTTACTCATCCACCCCATAGCCCGGCCAGTCTCTCCCGCCCGCTC	1380
DB	1403	AATGGAGGTGAACCTTACTCATCCACCCCATAGCCCGGCCAGTCTCTCCCGCCCGCTC	1462
QY	1381	ATCAGTCAGCGGTTGTGCACCAACAAACAGGCTGCGCCAGGCTTTTATCGSACCAAGCTT	1440
DB	1463	ATCAGTCAGCGGTTGTGCACCAACAAACAGGCTGCGCCAGGCTTTTATCGSACCAAGCTT	1522
QY	1441	CCCTCTCACATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCAATGAAACACAG	1500
DB	1523	CCCTCTCACATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCAATGAAACACAG	1588
QY	1501	CCAGCAGTTCCATGTGCGAGTCTCTAAGGGAAATTCAGATGTCAAACAGGCGTAGTCTCTGT	1560
DB	1583	CCAGCAGTTCCATGTGCGAGTCTCTAAGGGAAATTCAGATGTCAAACAGGCGTAGTCTCTGT	1642
QY	1561	AATGCTTTCAGCTTCTGTCCAAAACCTGCTGATGTTAATAGTCTCTCAGTGATCCACAGAACT	1620
DB	1643	AATGCTTTCAGCTTCTGTCCAAAACCTGCTGATGTTAATAGTCTCTCAGTGATCCACAGAACT	1702
QY	1621	CCTAAGAAACAAAAATTTACATCAGTATTCAACAGTTGCTGTTCCAGCAGTGTCAG	1680
DB	1703	CCTAAGAAACAAAAATTTACATCAGTATTCAACAGTTGCTGTTCCAGCAGTGTCAG	1762
QY	1681	TCTCAACCTAACCTTCATAGTAAATCTTTTGAGAAACCTTACCAAGCCGTTCCCTCTTCT	1740
DB	1763	TCTCAACCTAACCTTCATAGTAAATCTTTTGAGAAACCTTACCAAGCCGTTCCCTCTTCT	1822
QY	1741	ACCAATTACCAATTTCTGCAGTACAGTCTACCTTCGAACGCAATCTACGATGTCAGTTTCTAGT	1800
DB	1823	ACCAATTACCAATTTCTGCAGTACAGTCTACCTTCGAACGCAATCTACGATGTCAGTTTCTAGT	1882
QY	1801	AAAGTAAACAAACCGATCCCCCGCAGTGAAATCTGTCTCCAGCCCGTGATGTAATGCAAA	1860
DB	1883	AAAGTAAACAAACCGATCCCCCGCAGTGAAATCTGTCTCCAGCCCGTGATGTAATGCAAA	1942
QY	1861	TCCAGCTGAACTCCAGCGTCTGCTGCTCTATGGCGCCGAGTCCCTCTGAGGACTCTGAC	1920
DB	1943	TCCAGCTGAACTCCAGCGTCTGCTGCTCTATGGCGCCGAGTCCCTCTGAGGACTCTGAC	2002
QY	1921	GAGGAGTCAAAGGGGCTGGGCAAGGAGAAATGGGATTCGTACGATTTGAGCTCCCACTCT	1980
DB	2003	GAGGAGTCAAAGGGGCTGGGCAAGGAGAAATGGGATTCGTACGATTTGAGCTCCCACTCT	2062
QY	1981	CCCGGCCAAGATGCGGAAGATGAGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACC	2040
DB	2063	CCCGGCCAAGATGCGGAAGATGAGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACC	2122
QY	2041	CTAATAAGGTGCTAATAGTGCAGACAGCGACAGTGCACCGCAAGAAAAACGGCTAGCGCCT	2100
DB	2123	CTAATAAGGTGCTAATAGTGCAGACAGCGACAGTGCACCGCAAGAAAAACGGCTAGCGCCT	2182
QY	2101	GATGTGTCAGCTGCCAAGGCCAGCTCCCTCTGCACTCAGAAAAATCCCTTTGTAAAGGCA	2160
DB	2183	GATGTGTCAGCTGCCAAGGCCAGCTCCCTCTGCACTCAGAAAAATCCCTTTGTAAAGGCA	2242
QY	2161	AACGGTCTTCCTGGAAAGTTGATGCTGCTCTTGTGCTCTCTCCAGGAAGACAAAATC	2220
DB	2243	AACGGTCTTCCTGGAAAGTTGATGCTGCTCTTGTGCTCTCTCCAGGAAGACAAAATC	2302
QY	2221	TTAGAGACCTTCAGGCTTAGCAACAAACTGAAAGGCTCGACGGATGAAATCAGTGCACT	2280
DB	2303	TTAGAGACCTTCAGGCTTAGCAACAAACTGAAAGGCTCGACGGATGAAATCAGTGCACT	2362
QY	2281	GGAGCAGAGAGGGGGCCCTCCCGAGGACCGCAGCGCGAGCTTACGCTTGGAGCCCGCC	2340

2363 GGAGCAGAGAGGGGCTCTCCGAGGACCGGACGCGGAGCCTCAGCTGCGAGCCCGCC 2422
2341 GCCGAATCCCTCGAGGAGCCAGATGCGGCGCGCAAGCTT 2378
2423 GCCGAATCCCTCGAGGAGCCAGATGCGGCGCGCGGCT 2460

RESULT 3
AX882263
LOCUS
DEFINITION Sequence 17168 from Patent EP1074617.
ACCESSION AX882263
VERSION AX882263.1 GI:40037087
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1
Ota, T., Isegai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
Primers for synthesising full-length cDNA and their use
Patent: EP 1074617-A 17168 07-FEB-2001;
Research Association for Biotechnology (JP)
Location/Qualifiers
1. 3660
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
68. >3658
/note="unnamed protein product"
/codon_start=1
/db_xref="GI:40037088"
/translation="WTIVDKASESDPSAYONQPGSGEAVSPGDMWDAGSASWGAVSSI
NDVSNHLSLGPVCAVYVSSSDPSKSPQKQDALGDGIAIPQKVLFPSEKICLKL
WQOHRVAGAGLQNLGNTCFANAAQLQCLTYTPIPLANYMLSEHSKCTHAEGFCMCTWC
AHITQALSNPDGIVTKPMFVINEMRKRIARFNGQEDAEHFLQYTDVMAKACLNGSN
KLDRHTCATLTVQCI FGQYRSRMRKICNKGVSQTDFFDLQITLLEIKAAQSVNKALSE
PVKQIDIDGNSYKCSCKMVPASKRPTIHRSNVLTLSLKRFANFTGGKIAKDVK
FYKIDIFPMQSPNGEPIYVLYALVHTGFNCAGHYFCYIKASNLGWLQMDPSVAV
TSDIRSLTSCQAAYLVLYIRSHDVKNGBELTHPTSPGSSPRPVISRVVVTNKQAAP
FIGQLTSHMIKNPPLNGTGLKLPDTPSSMSPNGSSVNRASVPNASVQWVMSVAVS
RSSVTPHPKQKILTIISHNLKPVQCOSQNLHNSLENPTKVPSPSTITNSAVQS
SNATMSVSSKVTPIPRSESCSQVPMGKSLNGANSVLVPYGCSSSESDSESKLGL
ENGTGTVSSHSXPGDADEEATPHEIQEPTTLNGANSADSDSPKENGFLADPGACSC
GQALHNSNPFKANGLPEKLMKPAPLLSPEDKILLETFLNSLKLKSTDEMEAWAVAL
GPEDPRDAEPQPGSAAESLEPPAAGLSKTKAAPPRDPPTPTATKEGAWEMAAVAL
PEPPSAGEDIVGTPADPLDCGLSTGDASPLQDAKMAIAGEPRDSDALAEAPGLG
PAPPARSEPEQPLLPHPSGDHARDADQPSQSLGAPAEAREPPAPVLDMAPAHPEPS
DAEPPSGEEDAAAPKAPGPSAKELGSLKVDYRGHVRSPRSPSSSGSEPARSEVSI
TEGHRHRRTTCFPRERDQDRHAPHFGHGDRLSFGERRSLGRCSHHHSRHSRSL
DWRHHVTEGWRGKREXFPYDPRWDRCRYHDYLAARDMKFPFHGGRHERHAGC
HSDPHHNGRRCCEPARERHNSPSPRAGAPHALAPHDFDRFSDHRTALVAGONCI
1.SDRPHHNGRRCRSHSDVSNSSHVETKRAKSEQKDPLEEPKAKKHK"

ORIGIN

	Query Match	98.1%	Score 2365;	DB 6;	Length 3660;
	Best Local Similarity	99.8%;	Pred. No. 0;		
	Matches 2368;	Conservative 0;	Mismatches 5;	Indels 0;	Gaps 0;
Qy	6	GTGTGGCGGCGGCACGCCGCGCGAGGGGGATGGAGCGAGCCGCGAGCGCGGTTCAGAG	65		
Dd	1	GTGTGGCGGCGGCACGCCGCGCGAGGGGGATGGAGCGAGCCGCGAGCGCGGTTCAGAG	60		
Qy	66	TTGAACAATGACCATTGTCACAAAGCTTCTGAATCTTTCAGACCCCATCAGCCTTATCAGAA	125		
Dd	61	TTGACCAATGACCATTGTCACAAAGCTTCTGAATCTTTCAGACCCCATCAGCCTTATCAGAA	120		
Qy	126	TCAACCTGGCAGCTCCGAGGAGTCTCACCTGGAGACATGGATGACGGTCTTGCCAGCTG	185		
Dd	121	TCAACCTGGCAGCTCCGAGGAGTCTCACCTGGAGACATGGATGACGGTCTTGCCAGCTG	180		

186 GGGTCTGTGCTTCAATGAATGATGTGCTCAAAATCACACACTTTCTTTTAGGACCAATACC 245
181 GGGTCTGTGCTTCAATGAATGATGTGCTCAAAATCACACACTTTCTTTTAGGACCAATACC 240
246 TGGTCTGTAGTTTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACACAAAA 305
241 TGGTCTGTAGTTTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACACAAAA 300
306 GGATCAAGCCCTAGGTGATGTCATGCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAA 365
301 GGATCAAGCCCTAGGTGATGTCATGCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAA 360
366 GATTGTCTTAAGTGGCAACAAATCTATAGAGTTGGAGCTGGGTCCAGAAATTTGGGCAA 425
361 GATTGTCTTAAGTGGCAACAAATCTATAGAGTTGGAGCTGGGTCCAGAAATTTGGGCAA 420
426 TACCTGTTTTGCCAATGCAGCACTGCAGTGTGTTAACTTACACACACCTCTTGGCAATTA 485
421 TACCTGTTTTGCCAATGCAGCACTGCAGTGTGTTAACTTACACACACCTCTTGGCAATTA 480
486 CATGCTATCATGAACACACTCCAAAAATGTCATGCGAGAGGCTTTTGTATGATGTATC 545
481 CATGCTATCATGAACACACTCCAAAAATGTCATGCGAGAGGCTTTTGTATGATGTATC 540
546 AATGCAAGCACATATTACCCAGGCACCTCAGTAATCTCTGGGGACGTTATTAACCAATGTT 605
541 AATGCAAGCACATATTACCCAGGCACCTCAGTAATCTCTGGGGACGTTATTAACCAATGTT 600
606 TGTCAATCAATGAGATCGGGGTATAGTAGGCACTCCGTTTTTGGAAACCAAGAAATGC 665
601 TGTCAATCAATGAGATCGGGGTATAGTAGGCACTCCGTTTTTGGAAACCAAGAAATGC 660
666 CCATGAATTCCTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGCGACAA 725
661 CCATGAATTCCTCAATACACTGTTGATGCTATGCGAAGAGCATGCTTGAATGCGACAA 720
726 TAAATAGACAGACACACCAGGCCACCACTCTGTTTGTGTCAGTATTTGGAGATACCT 785
721 TAAATAGACAGACACACCAGGCCACCACTCTGTTTGTGTCAGTATTTGGAGATACCT 780
786 AAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGTTTCAGATACTTTTGATCCATATCT 845
781 AAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGTTTCAGATACTTTTGATCCATATCT 840
846 TGATATAACATTTGAGATAAAGCTGCTCAGAGTGTCAAAGGCATTTGGAGCAATTTGT 905
841 TGATATAACATTTGAGATAAAGCTGCTCAGAGTGTCAAAGGCATTTGGAGCAATTTGT 900
906 GAAGCCGAAACAGCTTGATGAGAAACTCGTCAAGTGCAGCAAGGTAAAGATGTT 965
901 GAAGCCGAAACAGCTTGATGAGAAACTCGTCAAGTGCAGCAAGGTAAAGATGTT 960
966 TCAGCTTCAAGAGGTTCACTATCCATAGATCCTTAATGTTCTTACACTTTCTCTGAA 1025
961 TCAGCTTCAAGAGGTTCACTATCCATAGATCCTTAATGTTCTTACACTTTCTCTGAA 1020
1026 ACSTTTTGAAATTTTACCGGTGAAATAATGCTTAAGATGTGAAATPACCTGAGTATCT 1085
1021 ACSTTTTGAAATTTTACCGGTGAAATAATGCTTAAGATGTGAAATPACCTGAGTATCT 1080
1086 TGATATTTGSCCATATATGTTCTCAACCAAGGAGGCCAATTTGTCACGTCCTGTATGC 1145
1081 TGATATTTGSCCATATATGTTCTCAACCAAGGAGGCCAATTTGTCACGTCCTGTATGC 1140
1146 AGTGTGCTCCACACTGGTTTTTAATGTCATGTCGCAATTACTTCTGCTACATAAAGC 1205
1141 AGTGTGCTCCACACTGGTTTTTAATGTCATGTCGCAATTACTTCTGCTACATAAAGC 1200
1206 TAGCAATGGCTCTGGTATCAAAATGAATGACTCCAATGTATCTACCAAGTATTTAGATC 1265
1201 TAGCAATGGCTCTGGTATCAAAATGAATGACTCCAATGTATCTACCAAGTATTTAGATC 1260

1266 GGTACTCAGCCAAACAAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG 1325
1261 GGTACTCAGCCAAACAAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG 1320
1326 AGGTGAATCTACTCATCCCACTAGCCCGGCGAGTCTCTCCCGCCCGTCATCAG 1385
1321 AGGTGAATCTACTCATCCCACTAGCCCGGCGAGTCTCTCCCGCCCGTCATCAG 1380
1386 TCAGCGGTTGTCAACCAACAAGGCTGCGCCAGGCTTTATTCGGACCAACAGCTTCCCTC 1445
1381 TCAGCGGTTGTCAACCAACAAGGCTGCGCCAGGCTTTATTCGGACCAACAGCTTCCCTC 1440
1446 TCACATGATAAAGAAATCCACTCATTAATGGAAGTGGACATTTGAAAGACACGCCAAG 1505
1441 TCACATGATAAAGAAATCCACTCATTAATGGAAGTGGACATTTGAAAGACACGCCAAG 1500
1506 CAGTTCATGTCAGTCTCTAACGGGAATTCAGAGTCAACAGGGCTAGTCTGTTAATGC 1565
1501 CAGTTCATGTCAGTCTCTAACGGGAATTCAGAGTCAACAGGGCTAGTCTGTTAATGC 1560
1566 TTCAGTCTCTGTCCAAACCTGGTCAGTTAATPAGTCTCTCAGTGTCCAGAACATCTCTAA 1625
1561 TTCAGTCTCTGTCCAAACCTGGTCAGTTAATPAGTCTCTCAGTGTCCAGAACATCTCTAA 1620
1626 GAAACAAAAATTACAAATCAGTATTCACAAAGTTCCTGTTCCGCCAGTGTCACTCA 1685
1621 GAAACAAAAATTACAAATCAGTATTCACAAAGTTCCTGTTCCGCCAGTGTCACTCA 1680
1686 ACCTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT 1745
1681 ACCTAACCTTCATAGTAATTTCTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT 1740
1746 TACCAATCTCGAGTACAGTCTACCTCGAAACGATCTACGATGTCACTTTCTAGTAAAGT 1805
1741 TACCAATCTCGAGTACAGTCTACCTCGAAACGATCTACGATGTCACTTTCTAGTAAAGT 1800
1806 AACAAACCGATCCCGCGAGTGAATCTGCTCCAGCCGCTGATGAATGGCAATCCAA 1865
1801 AACAAACCGATCCCGCGAGTGAATCTGCTCCAGCCGCTGATGAATGGCAATCCAA 1860
1866 GCTGAACTCCAGCGTGTGTCCTATGCGCGAGTCTCTGAGGACTCTGACGAGGA 1925
1861 GCTGAACTCCAGCGTGTGTCCTATGCGCGAGTCTCTGAGGACTCTGACGAGGA 1920
1926 GTCAAGGGGCTGGGCAAGGAGAAATGGATTTGTGATGTTGAGCTCCCACTCTCCCGG 1985
1921 GTCAAGGGGCTGGGCAAGGAGAAATGGATTTGTGATGTTGAGCTCCCACTCTCCCGG 1980
1986 CCAAGATCCGAAGATGAGAGGCCACTCCGCAAGAGCTTCAAGAACCCACGCCCTAAA 2045
1981 CCAAGATCCGAAGATGAGAGGCCACTCCGCAAGAGCTTCAAGAACCCACGCCCTAAA 2040
2046 CGGTGCTTAATAGTGCAGACAGCGACAGTGAACCGAAAGAAACGGCTAGCGCTGATGG 2105
2041 CGGTGCTTAATAGTGCAGACAGCGACAGTGAACCGAAAGAAACGGCTAGCGCTGATGG 2100
2106 TGCAGCTGCAAGGGCAGCTTGCCTTGCACTCAGAAAAATCCCTTTTGTAAAGCAACGG 2165
2101 TGCAGCTGCAAGGGCAGCTTGCCTTGCACTCAGAAAAATCCCTTTTGTAAAGCAACGG 2160
2166 TCTTCTGTGAAAGTTGATGCTGCTCTTTGCTGTCTCTCCCAAGAAACAAAAATCTTAGA 2225
2161 TCTTCTGTGAAAGTTGATGCTGCTCTTTGCTGTCTCTCTCCCAAGAAACAAAAATCTTAGA 2220
2226 GACCTTCAGCTTGAACAACTGAAAGCTCGAGGATGAATGATGTCACCTGAGC 2285
2221 GACCTTCAGCTTGAACAACTGAAAGCTCGAGGATGAATGATGTCACCTGAGC 2280
2286 AGAGAGGGGCTCTCCGAGGACCGCGACCGCAGCTTCAGCTTGGCAGCCCGCCGCGA 2345
2281 AGAGAGGGGCTCTCCGAGGACCGCGACCGCAGCTTCAGCTTGGCAGCCCGCCGCGA 2340
2346 ATCCCTGAGGAGCCAGATGGCGCCCGCAAGCTT 2378

[illegible]

D _b	1141	AGTCTGGTCCACATCGGTTTTAAATGCCATGCTGGCCATTACTTCTGCTACATAAAAGC	1200
Q _y	1206	TAGCAATGGCCTCTGGTATCAAAATGAATGATCCATTTGATCTACCAAGTGAATATAGATC	1265
D _b	1201	TAGCAATGGCCTCTGGTATCAAAATGAATGATCCATTTGATCTACCAAGTGAATATAGATC	1260
Q _y	1266	GGTACTCAGCCACACAGCCTATGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1325
D _b	1261	GGTACTCAGCCACACAGCCTATGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1320
Q _y	1326	AGGTGAATCTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGTTCATCAG	1385
D _b	1321	AGGTGAATCTACTCATCCACCCATAGCCCGGCGAGTCTCTCCCGCCCGTTCATCAG	1380
Q _y	1386	TCAGCGGTTCTGCACCAAAACAGCTGCGCCAGGCTTTATCGGACCACAGTCTCCCTC	1445
D _b	1381	TCAGCGGTTCTGCACCAAAACAGCTGCGCCAGGCTTTATCGGACCACAGTCTCCCTC	1440
Q _y	1446	TCACATGATAAGANTCCACCTCACTTAAATGGGACTGGACCATTTGAAAGACACGCCAAG	1505
D _b	1441	TCACATGATAAGANTCCACCTCACTTAAATGGGACTGGACCATTTGAAAGACACGCCAAG	1500
Q _y	1506	CAGTTTCCATGTCGAGTCTCTAAACGGGAAATCCAGTGTCAACAGGGCTAGTCTCTTAATGC	1565
D _b	1501	CAGTTTCCATGTCGAGTCTCTAAACGGGAAATCCAGTGTCAACAGGGCTAGTCTCTTAATGC	1560
Q _y	1566	TTTCAAGTTCTGTGCCAAAACCTGTGTCAATAGTGTCTCAGTGATCCAGAACATCCTAA	1625
D _b	1561	TTTCAAGTTCTGTGCCAAAACCTGTGTCAATAGTGTCTCAGTGATCCAGAACATCCTAA	1620
Q _y	1626	GAAACAAAATAATCAATCAGTATTCACACAAAGTTGCGCTTTCGGCAGTGTCAAGTCTCA	1685
D _b	1621	GAAACAAAATAATCAATCAGTATTCACACAAAGTTGCGCTTTCGGCAGTGTCAAGTCTCA	1680
Q _y	1686	ACCTAACCTTCATAGTAAATCTTTGGAGAACCCCTACCAAGCCCGTTCCTCTCTACCAT	1745
D _b	1681	ACCTAACCTTCATAGTAAATCTTTGGAGAACCCCTACCAAGCCCGTTCCTCTCTACCAT	1740
Q _y	1746	TACCAATTTCTGCASTACAGTCTACCTCGAACCGCATCTACGATGTCAAGTTCTAGTAAAGT	1805
D _b	1741	TACCAATTTCTGCASTACAGTCTACCTCGAACCGCATCTACGATGTCAAGTTCTAGTAAAGT	1800
Q _y	1806	AACAAAACCGATCCCGGAGTGAATCTCTGCCAGCCGCTGATGAATGGCAAAATCCAA	1865
D _b	1801	AACAAAACCGATCCCGGAGTGAATCTCTGCCAGCCGCTGATGAATGGCAAAATCCAA	1860
Q _y	1866	GCTGAATCTCAGCGTCTGCTGCTGCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGA	1925
D _b	1861	GCTGAATCTCAGCGTCTGCTGCTGCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGA	1920
Q _y	1926	GTCAAAAGGGGTGGCAAGAGAGATGGGATTTGGTACGATTTGTGAGCTCCCACTCCCGG	1985
D _b	1921	GTCAAAAGGGGTGGCAAGAGAGATGGGATTTGGTACGATTTGTGAGCTCCCACTCCCGG	1980
Q _y	1986	CCAAAGATCCGAAAGATGAGAGGGCCACTCCGACAGGCTTCAAGAACCCATGACCTTAAA	2045
D _b	1981	CCAAAGATCCGAAAGATGAGAGGGCCACTCCGACAGGCTTCAAGAACCCATGACCTTAAA	2040
Q _y	2046	CGGTGCTTAATAGTGCAGACACGACAGTACCCGAAAGAAAACGGCTTAGCGCTGATGG	2105
D _b	2041	CGGTGCTTAATAGTGCAGACACGACAGTACCCGAAAGAAAACGGCTTAGCGCTGATGG	2100
Q _y	2106	TGCCAGCTGCCAAGGCCACCGCTGCCCTGCATCTAGAAAAATCCCTTTGTAGGGCAACGG	2165
D _b	2101	TGCCAGCTGCCAAGGCCACCGCTGCCCTGCATCTAGAAAAATCCCTTTGTAGGGCAACGG	2160
Q _y	2166	TCTTCTCGAAAGTTGATGCTGCTCTCTTGTCTCTCCAGAGACAAAATCTTAGA	2225
D _b	2161	TCTTCTCGAAAGTTGATGCTGCTCTCTTGTCTCTCTCCAGAGACAAAATCTTAGA	2220
Q _y	2226	GACCTTCAGGCTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGCACTGGAGC	2285
D _b	2221	GACCTTCAGGCTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGCACTGGAGC	2280

QY	2286	AGAGAGGGGGCTCCCGAGGACCGCGAGCGCTCAGCTGGCAGCCCGCGCCGA	2340
Db	2281	AGAGAGGGGGCTCCCGAGGACCGCGAGCGCTCAGCTGGCAGCCCGCGCCGA	2340
QY	2346	ATCCCTGGAGGACGACATGGCGCGCAAGCTT	2378
Db	2341	ATCCCTGGAGGACGACATGGCGCGCGCT	2373
RESULT 6			
LOCUS	BC060846	3669 bp	linear
DEFINITION	Homo sapiens ubiquitin specific protease 42, mRNA (cdna clone IMAGE:30343487), partial cds.		
ACCESSION	BC060846		
VERSION	BC060846.1 GI:38173811		
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 3669)		
AUTHORS	Strausberg,R.D., Collins,F.S., Wagner,L., Shemen,C.M., Schuler,G.D., Klausner,R.D., Zeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,S.F., Jordan,K., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,K., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Donald,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaby,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S.D., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shvchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalley,D.E., Schenker,A., Schein,J.E., Jones,S.J. and Marra,M.A.		
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)		
MEDLINE	22388257		
PUBMED	12477923		
REFERENCE	2 (bases 1 to 3669)		
AUTHORS	Strausberg,R.		
TITLE	Direct Submission		
JOURNAL	Submitted (03-NOV-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs@mail.nih.gov Tissue Procurement: Dr. Stefan Hansson cDNA Library Preparation: Michael Brownstein / Ted Usdin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www.shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.		

QY	306	GGATCAAGCCCTAGGTTGATGGCATCGCTCCTCCACAGAAAGTCTCTTTTCCCATCTGAGAA	365
DB	296	GGATCAAGCCCTAGGTTGATGGCATCGCTCCTCCACAGAAAGTCTCTTTTCCCATCTGAGAA	355
QY	366	GATTGTGTTTAAAGTGGCAACAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGGCAG	425
DB	356	GATTGTGTTTAAAGTGGCAACAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGGCAG	415
QY	426	TACCTGTTTGGCCAAATGCAGCACTGCAGTGTTTAACTTACACACACACCTCTTGCCCAATTA	485
DB	416	TACCTGTTTGGCCAAATGCAGCACTGCAGTGTTTAACTTACACACACCTCTTGCCCAATTA	475
QY	486	CATGCTATCACATGAACCTTCCAAACATCTCATGAGAGGCTTTTGTATGATGTGATC	545
DB	476	CATGCTATCACATGAACCTTCCAAACATCTCATGAGAGGCTTTTGTATGATGTGATC	535
QY	546	AATGCAAGCACATATTACCAGGCACCTCAGTAATTCCTGGGAGCTTATTAAACCAATGTT	605
DB	536	AATGCAAGCACATATTACCAGGCACCTCAGTAATTCCTGGGAGCTTATTAAACCAATGTT	595
QY	606	TGTCATCAATGAGATGGGGGTATAGCTAGGCACCTCCGTTTGGGAAACCAAGAGATGC	665
DB	596	TGTCATCAATGAGATGGGGGTATAGCTAGGCACCTCCGTTTGGGAAACCAAGAGATGC	655
QY	666	CCATGAATTCCTTCAATACACTGTGTGATGCTATGCAAGAAACGATGTTGAATGGCAGCAA	725
DB	656	CCATGAATTCCTTCAATACACTGTGTGATGCTATGCAAGAAACGATGTTGAATGGCAGCAA	715
QY	726	TAAATTTAGACAGACACACCCAGGCCACACTCTTGTGTTTGCAGATATTTGGAGGATACCT	785
DB	716	TAAATTTAGACAGACACACCCAGGCCACACTCTTGTGTTTGCAGATATTTGGAGGATACCT	775
QY	786	AGATCTAGATCAAAATGTTTAAATTCGACGGGGCTTTCAGATACCTTTTGATCCATATCT	845
DB	776	AGATCTAGATCAAAATGTTTAAATTCGACGGGGCTTTCAGATACCTTTTGATCCATATCT	835
QY	846	TCATATAACATTTGGAGATAAAGGGTGTCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGT	905
DB	836	TCATATAACATTTGGAGATAAAGGGTGTCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGT	895
QY	906	GAAGCCGGAAACAGCTTTGATGGAGAAACCTCGTACAGTGCAGCAAGTGTAAAAGATGTT	965
DB	896	GAAGCCGGAAACAGCTTTGATGGAGAAACCTCGTACAGTGCAGCAAGTGTAAAAGATGTT	955
QY	966	TCCAGGTTCAAGAGAGGTTCACTATCCATAGATCCCTAAATGTTCTTACACTTCTCTGAA	1025
DB	956	TCCAGGTTCAAGAGAGGTTCACTATCCATAGATCCCTAAATGTTCTTACACTTCTCTGAA	1015
QY	1026	AGCTTTTGCAAAATTTTACCGTGGAAAAATTTGCTAAGGAATGTGAAATACCTCGAGTATCT	1085
DB	1016	AGCTTTTGCAAAATTTTACCGTGGAAAAATTTGCTAAGGAATGTGAAATACCTCGAGTATCT	1075
QY	1086	TGATATTCGGCCATATATGTTCTCAACCCAAACGAGAGCCATTTGCTACGCTCTGTATGC	1145
DB	1076	TGATATTCGGCCATATATGTTCTCAACCCAAACGAGAGCCATTTGCTACGCTCTGTATGC	1135
QY	1146	AGTGCTGGTCCACACTGGTTTTTAATTTGCCATGCTGGCCATTTACTTCTGCTACATAAAAGC	1205
DB	1136	AGTGCTGGTCCACACTGGTTTTTAATTTGCCATGCTGGCCATTTACTTCTGCTACATAAAAGC	1195
QY	1206	TAGCAATGGCCCTCTGGTATCAAAATGAATGATCTCCATTTGTATCTACCAAGTGATTTAGATC	1265
DB	1196	TAGCAATGGCCCTCTGGTATCAAAATGAATGATCTCCATTTGTATCTACCAAGTGATTTAGATC	1255
QY	1266	GGTACTCAGCCCAACAGCCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1325
DB	1256	GGTACTCAGCCCAACAGCCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1315
QY	1326	AGGTGAATTTACTTCATCCACCATAGCCCCCGCCAGCTCTCTCCCGCCCGCCGTCATCAG	1385
DB	1316	AGGTGAATTTACTTCATCCACCATAGCCCCCGCCAGCTCTCTCCCGCCCGCCGTCATCAG	1375
QY	1386	TCAGCGGGTGTGCCAAACAAACAGGCGTGGCCAGGCTTTTATCGAACCAAGAGCTTCCCTC	1445

Db	1376	TCACGGGGTTGTCAACCAACAAACAGCGCTGGCCAGGCTTTATCGGACACACAGCTTCCCTC	1435
Qy	1446	TCACATGATAAAGAAATCCACCTCACCTTAAATGGGACTGGACCAATTTGAAGAAGACACGCCAAG	1505
Db	1436	TCACATGATAAAGAAATCCACCTCACCTTAAATGGGACTGGACCAATTTGAAGAAGACACGCCAAG	1495
Qy	1506	CAGTTCCATGTCGAGTCTCAACGGGAATTCGAGTGTCAACAGGGCTAGTCCGTGTTAAATGC	1565
Db	1496	CAGTTCCATGTCGAGTCTCAACGGGAATTCGAGTGTCAACAGGGCTAGTCCGTGTTAAATGC	1555
Qy	1566	TTACAGCTTCTGTCTCAAAACTGGTCAGTTAATAGTCTCCTCAGTGTATCCAGAACATCTCTAA	1625
Db	1556	TTACAGCTTCTGTCTCAAAACTGGTCAGTTAATAGTCTCCTCAGTGTATCCAGAACATCTCTAA	1615
Qy	1626	GAACACAAAAATTACAAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCACTCTCA	1685
Db	1616	GAACACAAAAATTACAAATCAGTATTCACAAAGTTGCTGTTCGCCAGTGTCACTCTCA	1675
Qy	1696	ACCTAAACCTTCATAGTAATCTTTTGGAGAACCCCTACCAAGCCCGTTCCCTCTTCTACCAT	1745
Db	1676	ACCTAAACCTTCATAGTAATCTTTTGGAGAACCCCTACCAAGCCCGTTCCCTCTTCTACCAT	1735
Qy	1746	TACCAATTTCTGCAGTACAGTCTACCTCGAAACGATCTACGATGTGAGTTCTAGTAAAGT	1805
Db	1736	TACCAATTTCTGCAGTACAGTCTACCTCGAAACGATCTACGATGTGAGTTCTAGTAAAGT	1795
Qy	1806	AACAAACCGATCCCGCGAGTGAATCCTGCTCCAGCCCGTCAATGAATGGCAAAATCCAA	1865
Db	1796	AACAAACCGATCCCGCGAGTGAATCCTGCTCCAGCCCGTCAATGAATGGCAAAATCCAA	1855
Qy	1866	GCTGAACCTCAGCGCTGTGCTGCTGCTATGGCGCGAGTCTCTTGAAGACTCTCAACGAGGA	1925
Db	1856	GCTGAACCTCAGCGCTGTGCTGCTGCTATGGCGCGAGTCTCTTGAAGACTCTCAACGAGGA	1915
Qy	1926	GTCAAAGGGCTGGCAAGAGAAATGGGATTTGGTACGATTTGTGAGTCCCACTCTCCCG	1985
Db	1916	GTCAAAGGGCTGGCAAGAGAAATGGGATTTGGTACGATTTGTGAGTCCCACTCTCCCG	1975
Qy	1986	CCAAGATCCGAAGATGAGGAGGCCACTCCGACAGCTTCAAGAACCCATGACCCCTAA	2045
Db	1976	CCAAGATCCGAAGATGAGGAGGCCACTCCGACAGCTTCAAGAACCCATGACCCCTAA	2035
Qy	2046	CGGTGCTAATAGTTCAGACAGCGACAGTACCCGAAAGAAACCGGCTAGCGCTGATGG	2105
Db	2036	CGGTGCTAATAGTTCAGACAGCGACAGTACCCGAAAGAAACCGGCTAGCGCTGATGG	2095
Qy	2106	TGCGAGCTGCCAAAGCGACGCTGCCCTGCATCTCAGAAATCCCTTTGCTAAGGCACACGG	2165
Db	2096	TGCGAGCTGCCAAAGCGACGCTGCCCTGCATCTCAGAAATCCCTTTGCTAAGGCACACGG	2155
Qy	2166	TCTTCTCGAAAGTTGATGCTGCTCCTTGTCTCTCCAGAAACAAATCTTTAGA	2225
Db	2156	TCTTCTCGAAAGTTGATGCTGCTCCTTGTCTCTCCAGAAACAAATCTTTAGA	2215
Qy	2226	GACCTTCAGGCTTAGCAACAACTGAAGGCTCGACCGATGAATGATGACCTGGAGC	2285
Db	2216	GACCTTCAGGCTTAGCAACAACTGAAGGCTCGACCGATGAATGATGACCTGGAGC	2275
Qy	2286	AGAGAGGGGCCCTCCGAGGACCGGACCGGAGCCTCAGCGCTGGCAGCCCCCGCGCGA	2345
Db	2276	AGAGAGGGGCCCTCCGAGGACCGGACCGGAGCCTCAGCGCTGGCAGCCCCCGCGCGA	2335
Qy	2346	ATCCCTGGAGGAGCCAGATGCGGCGCAAGCTT	2378
Db	2336	ATCCCTGGAGGAGCCAGATGCGGCGCGGCGCT	2368

RESULT 7
AX107851
LOCUS
DEFINITION
ACCESSION

AX107851	AX107851	2347 bp
LOCUS	Sequence 2 from Patent WO0123589.	
DEFINITION		
ACCESSION	AX107851	


```
VERSION AX107851.1 GI:13923242
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL Kapeller-Libermann, R.
TITLE 23431, a novel human ubiquitin protease
JOURNAL Patent: WO 0123589-A 2 05-APR-2001;
MILLennium Pharmaceuticals, Inc. (US)
FEATURES Location/Qualifiers
source 1..2347
         /organism="Homo sapiens"
         /mol_type="unassigned DNA"
         /db_xref="taxon:9606"
         /note="23431 Ubiquitin protease"
         61..>2347
         /note="unnamed protein product"
         /codon_start=1
         /protein_id="CAC37760.1"
         /db_xref="GI:13923243"
         /db_xref="REVIEW:13923243"
         /translation="MTIVDKASSESDPSAYIQNPGSSSEAVSPDMDAGSASWGVSSL
NDVSHNTSLGVPVGAIVYSSVDPKSPQKQALGDGIAPKPVLPPEKIKLK
WQTHRVGAGLQNLGNTCFANAALQCLTYTPPLANTMLSHEHSKTHARFGCMWQ
AHIQALSNPQVLIKPMFVFNMRRIARHFRGNODEHEFLQYTVDMQKACLNGSN
KLDRTQATLTCVQIFGGVLSRVKCLNCKGVSDPFDYLDLLEIKAAQSVKALEQ
FVKPQLDGENSYKSKCKMVPASRFTIHRSSNVLTLKRFANFTGGKIAQDVKY
PEYLDIRPSPNGPEPIVYLVAVLHVFNFCHAGHYFYIKASNGLWYQWDSIVS
TSDIRSVLSQQAVLYFIRSHDVXNGELTHPSQSPRPVLSQVVTNKQAAPG
FIGPQLPMKINPPLHNGTGLKDTPPSSMSPNCSVNRASVNASVQNMNVN
RSVLPHPKPKQKTIITSHNKLVPQCSQPNLHNSLENPTKVPSPITINSVQST
SNASVMSVSKYTKPIPRSESCSQPMWKGKLSVLVPGAESDSEDESKGLGK
ENIGITIVSHSPQDADEATPHELPQPMTLNGANSADSDPKENGLAPDGASCQ
GPPALHNSNPFKANGLPKMPALPILLSLPEDKILETLFRLSNKLKSGTDEMSAPGAER
GPPEDRDABPQPSPAESLES"
ORIGIN
Query Match 97.0%; Score 2339.2; DB 6; Length 2347;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2341; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 16 GCGGGCGGGCGGGCGGAGGAGTGAGCGAGCGCGCGAGCGGGTCAGAGTTGAACAATG 75
DB 4 GCGTCGCGGGCGGGCGGAGGAGTGAGCGAGCGCGCGGTCAGAGTTGAACAATG 63
QY 76 ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAATCAGCCTGGC 135
DB 64 ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAATCAGCCTGGC 123
QY 136 AGCTCCGAGGAGTCTACCTGGAGACATGGATGAGGTTCTGCGAGCTGGGGTCTGTG 195
DB 124 AGCTCCGAGGAGTCTACCTGGAGACATGGATGAGGTTCTGCGAGCTGGGGTCTGTG 183
QY 196 TCTTCATTGAATGATGTGCAAAATCACACACTTCTTTAGGACCACTACCTGCTGCTGA 255
DB 184 TCTTCATTGAATGATGTGCAAAATCACACACTTCTTTAGGACCACTACCTGCTGCTGA 243
QY 256 GTTTATTGAGTTGATCTGATCAATCAAAATCAAAACCATCACCACAAAGGATCAAGCC 315
DB 244 GTTTATTGAGTTGATCTGATCAATCAAAATCAAAACCATCACCACAAAGGATCAAGCC 303
QY 316 CTAGTGATGATCGCTCCCTCCACAGAAAGTTCTTTCCCATCTGAGAAAGTTGTCTT 375
DB 304 CTAGTGATGATCGCTCCCTCCACAGAAAGTTCTTTCCCATCTGAGAAAGTTGTCTT 363
QY 376 AAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTT 435
DB 364 AAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTT 423
QY 436 GCCAATGAGCACTGACGTGTTTAACTTACACACACCTCTTGGCAATTAACATGCTATCA 495
```

QY	1576	GTCCAAAAC	CTGGTCA	CAGTTAA	TAGTCTC	CAGTGAT	CCCGA	ACATCT	TAAGAAA	CAAAA	1635
Db	1564	GTCCAAAAC	TGGTCA	GTTAA	TAGTCTC	CAGTGAT	CCCGA	ACATCT	TAAGAAA	CAAAA	1623
QY	1636	ATTACAA	TTCAGTAT	TACAA	CAAGTTG	CGCTGTT	CGCCAG	TGTCA	GTCACT	CAACCTA	1695
Db	1624	ATTACAA	TTCAGTAT	TACAA	CAAGTTG	CGCTGTT	CGCCAG	TGTCA	GTCACT	CAACCTA	1683
QY	1696	CATAGTAA	TTCTTTG	GAGAAC	CCCTAC	CAAGCCGG	TTCCTCT	TTCAC	CAATAC	CAATTC	1755
Db	1684	CATAGTAA	TTCTTTG	GAGAAC	CCCTAC	CAAGCCGG	TTCCTCT	TTCAC	CAATAC	CAATTC	1743
QY	1756	GCAGTAC	AGTCTAC	CTCGA	ACGATCT	ACAGTGT	CAGTTT	CTAGTAA	AGTAA	CAAAA	1815
Db	1744	GCAGTAC	AGTCTAC	CTCGA	ACGATCT	ACAGTGT	CAGTTT	CTAGTAA	AGTAA	CAAAA	1803
QY	1816	ATCCCCG	CAGTGA	ATCTCT	CTCCAC	CGCGTG	ATGAAT	TGGCA	AAATCAA	AGTGA	1875
Db	1804	ATCCCCG	CAGTGA	ATCTCT	CTCCAC	CGCGTG	ATGAAT	TGGCA	AAATCAA	AGTGA	1863
QY	1876	AGCGTGT	GTGCGC	CTATG	CGCGC	AGTCTCT	CTGAGG	ACTCTG	ACGAG	CAGTCA	1935
Db	1864	AGCGTGT	GTGCGC	CTATG	CGCGC	AGTCTCT	CTGAGG	ACTCTG	ACGAG	CAGTCA	1923
QY	1936	CTGGCA	AAGCAG	AATGG	GAATGG	TATG	TGTAC	CAATTT	GTGAG	CTCCC	1995
Db	1924	CTGGCA	AAGCAG	AATGG	GAATGG	TATG	TGTAC	CAATTT	GTGAG	CTCCC	1983
QY	1996	GAAGAT	GAGGAG	GCAC	CTCCG	CACG	AGCTT	CAAG	ACCCAT	GACCTTAA	2055
Db	1984	GAAGAT	GAGGAG	GCAC	CTCCG	CACG	AGCTT	CAAG	ACCCAT	GACCTTAA	2043
QY	2056	AGTCACA	CAGCAG	CATG	ATCCG	GAAGAAA	CGGCTAG	CGCTCAT	GTGTC	CCAGCTGC	2115
Db	2044	AGTCACA	CAGCAG	CATG	ATCCG	GAAGAAA	CGGCTAG	CGCTCAT	GTGTC	CCAGCTGC	2103
QY	2116	CAAGGC	CAGCCT	GCCTG	CACTC	AGAAAA	TCCCTTT	GTCTA	AGGCA	AAACGGT	2175
Db	2104	CAAGGC	CAGCCT	GCCTG	CACTC	AGAAAA	TCCCTTT	GTCTA	AGGCA	AAACGGT	2163
QY	2176	AAGTTG	ATGCTGT	CTCTTT	GTCTCT	CCGAG	AGACAAA	ATCTT	TAG	AGACCTT	2235
Db	2164	AAGTTG	ATGCTGT	CTCTTT	GTCTCT	CCGAG	AGACAAA	ATCTT	TAG	AGACCTT	2223
QY	2236	CTTAGC	ACAACT	GAAGG	CTC	ACGAT	GAATG	ATG	ATG	ATG	2295
Db	2224	CTTAGC	ACAACT	GAAGG	CTC	ACGAT	GAATG	ATG	ATG	ATG	2283
QY	2296	CCTCCG	AGGAC	CGCGA	CGCC	TAG	CGCTC	AGC	CTG	GGCAG	2355
Db	2284	CCTCCG	AGGAC	CGCGA	CGCC	TAG	CGCTC	AGC	CTG	GGCAG	2343
QY	2356	GAGC	2359								
Db	2344	GAGC	2347								

RESULT 8			
AX364931			
LOCUS	AX364931	1679 bp	DNA linear
DEFINITION	Sequence 82 from Patent WO0206315.		

VERSION	AX364931.1	GI:18696821
KEYWORDS		
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	Mintz, L., Freilich, S. and Bernstein, J.	
AUTHORS	Novel nucleic acid and amino acid sequences	
TITLE	Patent: WO 0206315-A 82 24-JAN-2002;	
JOURNAL		

```
Db 901 AATAGTCTCTCAGTATCCAGAACATCTTAAGAAACAAAATTAACAATCAGTATTCAC 960
QY 1654 AACAGTGGCTCTTCCGCGAGTCTCAGTCTCAACCTTAACCTTATAGTAATCTTTGGAG 1713
Db 961 AACAGTGGCTCTTCCGCGAGTCTCAGTCTCAACCTTAACCTTATAGTAATCTTTGGAG 1020
QY 1714 AACCTTACCAAGCCGCTTCCCTCTTACCAATTAACCAATTCAGTACAGTCTCACTCG 1773
Db 1021 AACCTTACCAAGCCGCTTCCCTCTTACCAATTAACCAATTCAGTACAGTCTCACTCG 1080
QY 1774 AAGGATCTACGATGTCAGTCTTCTAGTAAGTAAACAAAACCCAGTCCCGCAGTGAATCC 1833
Db 1081 AAGCATCTACGATGTCAGTCTTCTAGTAAGTAAACAAAACCCAGTCCCGCAGTGAATCC 1140
QY 1834 TGCTCCAGCCGCTGATGTCGAATGTCGAATCAACCTCAAGCTGAACTCCAGCGTCTGTCGCTAT 1893
Db 1141 TGCTCCAGCCGCTGATGTCGAATGTCGAATCAACCTCAAGCTGAACTCCAGCGTCTGTCGCTAT 1200
QY 1894 GCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGAAGGAGGAGATGGG 1953
Db 1201 GCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGAAGGAGGAGATGGG 1260
QY 1954 ATTGTAAGATGTCAGTCTCCACTCTCCCGGCAAGATGCGGAAGATGAGGAGGCACT 2013
Db 1261 ATTGTAAGATGTCAGTCTCCACTCTCCCGGCAAGATGCGGAAGATGAGGAGGCACT 1320
QY 2014 CGCGACGAGTCTCAAGAACCCATGACCTTAACGGTGTCTAATGTGACAGACGACAGT 2073
Db 1321 CGCGACGAGTCTCAAGAACCCATGACCTTAACGGTGTCTAATGTGACAGACGACAGT 1380
QY 2074 GACCCGAAAGAAACGGCTAGCGCTGATGTGCGCAGCTGCAAGGCGCAGCTGCCCTG 2133
Db 1381 GACCCGAAAGAAACGGCTAGCGCTGATGTGCGCAGCTGCAAGGCGCAGCTGCCCTG 1440
QY 2134 CACTCAGAAATCCCTTGTCTAAGCAACCGTCTTCTCGAAAGTGTATGCTGCTCT 2193
Db 1441 CACTCAGAAATCCCTTGTCTAAGCAACCGTCTTCTCGAAAGTGTATGCTGCTCT 1500
QY 2194 TTGCTGTCTCTCCAGACAGCAAAATCTTAGACGCTTCAGGCTTAGCAAACTGAAA 2253
Db 1501 TTGCTGTCTCTCCAGACAGCAAAATCTTAGACGCTTCAGGCTTAGCAAACTGAAA 1560
QY 2254 GGCTCGACGATGAATGATGTCACCTTGAGCAGAGAGGGGCTCCGAGGACCGGAC 2313
Db 1561 GGCTCGACGATGAATGATGTCACCTTGAGCAGAGAGGGGCTCCGAGGACCGGAC 1620
QY 2314 GCGAGCTCTAGCTGGCAGCCCGCGCGAATCCCTGGAGGACCCAGATGCGGCGCG 2372
Db 1621 GCGAGCTCTAGCTGGCAGCCCGCGCGAATCCCTGGAGGACCCAGATGCGGCGCG 1679
```

```
RESULT 9
AC004895/c
LOCUS AC004895 152927 bp DNA linear PRI 30-SEP-2000
DEFINITION Homo sapiens PAC clone RP4-810E6 from 7, complete sequence.
ACCESSION AC004895
VERSION AC004895.2 GI:4926908
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152927)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
9063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 152927)
Cordes, M. and Gibson, A.
The sequence of Homo sapiens PAC clone RP4-810E6
Unpublished
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
```

```
3 (bases 1 to 152927)
Waterston, R.H.
Direct Submission
Submitted (12-JUN-1998) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 152927)
Waterston, R.H.
Direct Submission
Submitted (29-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 152927)
Waterston, R.
Direct Submission
Submitted (30-SEP-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 29, 1999 this sequence version replaced gi:3450909.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_DJ0810E06
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:regreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor. The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong. VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP1-42M2. Actual start of this clone is at base position 1 of RP4-810E6; actual end is at 152927 of RP4-810E6.

There is a questionable number of A's from 81574-81601. The consensus reflects 27 A's but could be as many as 29.

```
Location/Qualifiers
1. .152927
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
```

FEATURES
source

[illegible]

```

Db      46241  TCTGACGAGAGTCAAGGGGCTGGCAAGGAGATGGGATGGTACGATTGTGAGCTCC 46182
QY      1975  CACTCTCCCGGCAAGATGCGGAAGATGAGAGGCCACTCGACGAGCTTCAAGAACCC 2034
Db      46181  CACTCTCCCGGCAAGATGCGGAAGATGAGAGGCCACTCGACGAGCTTCAAGAACCC 46122
QY      2035  ATGACCTAAACGGTGTCTATAGTGCAGACAGCGACAGTACCCGGAAGAAAACGGCTA 2094
Db      46121  ATGACCTAAACGGTGTCTATAGTGCAGACAGCGACAGTACCCGGAAGAAAACGGCTA 46062
QY      2095  GCGCTGTATGTCAGCTGCCAAGCGCACCTGCCCTGCCTCAGTACAGAAAATCCCTTTGCT 2154
Db      46061  GCGCTGTATGTCAGCTGCCAAGCGCACCTGCCCTGCCTCAGTACAGAAAATCCCTTTGCT 46002
QY      2155  AAGGCAACGGTCTTCTCGAAGTTGA 2182
Db      46001  AAGGCAACGGTCTTCTCGAAGTTGA 45974

RESULT 10
AC069393
LOCUS   Homo sapiens chromosome 7 clone RP11-585K18, WORKING DRAFT
DEFINITION
AC069393
VERSION 1 (bases 1 to 154791)
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
           Waterston,R.H.
           The sequence of Homo sapiens clone
           Unpublished
           2 (bases 1 to 154791)
           Waterston,R.H.
           Direct Submission
           Submitted (28-MAY-2000) Genome Sequencing Center, Washington
           University School of Medicine, 4444 Forest Park Parkway, St. Louis,
           MO 63108, USA
           On Oct 26, 2000 this sequence version replaced gi:10946568.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H NH0585K18
----- Summary Statistics -----
Sequencing vector: M13; 71%
Chemistry: Dye-terminator Big Dye; 5% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 143345 bases at least Q40
Consensus quality: 146572 bases at least Q30
Consensus quality: 148386 bases at least Q20
Insert size: 164000; agarose-fp
Quality coverage: 4.65 in Q20 bases; agarose-fp
Quality coverage: 5.04 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
5210    5309: gap of unknown length
5310    10071: contig of 4762 bp in length
10072    10171: gap of unknown length
10172    16164: contig of 5993 bp in length
16165    16264: gap of unknown length
16265    29394: contig of 13130 bp in length
29395    29494: gap of unknown length
29495    41590: contig of 12096 bp in length
41591    41690: gap of unknown length
41691    54703: contig of 13013 bp in length
54704    54803: gap of unknown length
54804    74381: contig of 19578 bp in length
74382    74481: gap of unknown length
74482    103253: contig of 28772 bp in length
103254    103354: gap of unknown length
103355    139431: contig of 36078 bp in length
139432    139531: gap of unknown length
139532    140718: contig of 1187 bp in length
140719    140818: gap of unknown length
140819    143471: contig of 2653 bp in length
143472    143571: gap of unknown length
143572    147062: contig of 3491 bp in length
147063    147162: gap of unknown length
147163    150060: contig of 2898 bp in length
150061    150161: gap of unknown length
150162    154791: contig of 4631 bp in length.

Location/Qualifiers
1..154791
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-585K18"
1..5209
/note="assembly_name:Contig10"
5310..10071
/note="assembly_name:Contig11"
10172..16164
/note="assembly_name:Contig12"
clone_end:T7
vector_side:left"
16265..29394
/note="assembly_name:Contig13"
29495..41590
/note="assembly_name:Contig14"
41691..54703
/note="assembly_name:Contig15"
54804..74381
/note="assembly_name:Contig16"
74482..103253
/note="assembly_name:Contig17"
103354..139431
/note="assembly_name:Contig18"
139532..140718
/note="assembly_name:Contig19"
clone_end:SP6
vector_side:right"
140819..143471
/note="assembly_name:Contig6"
143572..147062
/note="assembly_name:Contig7"
147163..150060
/note="assembly_name:Contig8"
150161..154791
/note="assembly_name:Contig9"

FEATURES
source

```

ORIGIN

Query Match 30.1%; Score 727.2; DB 2; Length 154791;
 Best local Similarity 98.3%; Pred. No. 7.6e-197;
 Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1435 CAGCTTCCCTCTCAGATGATAAGATCCACCTCAGCTTAATGAGGACCATTCGAAA 1494
 ||||| 1 |||||

Db	129179	CAGGTGGCTCTGCTTACTTTCAGAAATCCACCTCACTTAAATGGAGCTGGACCATTTGAA	129238
QY	1495	GACAGCCAAAGCAGTTCATGTCGAGTCCCTAAACGGGAATTCAGTGTCAACAGGGCTAGT	1554
Db	129239	GACAGCCAAAGCAGTTCATGTCGAGTCCCTAAACGGGAATTCAGTGTCAACAGGGCTAGT	129298
QY	1555	CCTGTTAATGCTTCAGCTTCTGTCCTCAAAATCTGGTCAAGTAAATAGTCTCAGTATCCCA	1614
Db	129299	CCTGTTAATGCTTCAGCTTCTGTCCTCAAAATCTGGTCAAGTAAATAGTCTCAGTATCCCA	129358
QY	1615	GAACATCTTAAGAAACAAAATTAACATCAGTATTCAACAAGTTGCTGTTCGCCAG	1674
Db	129359	GAACATCTTAAGAAACAAAATTAACATCAGTATTCAACAAGTTGCTGTTCGCCAG	129418
QY	1675	TGTCACTCTCAACCTTCAACCTTCACTAGTAATCTTTTGAGAACCTTCAACAGCCCGTTCC	1734
Db	129419	TGTCACTCTCAACCTTCAACCTTCACTAGTAATCTTTTGAGAACCTTCAACAGCCCGTTCC	129478
QY	1735	TCCTTACCATTACCAATTCCTGAGTACGATCTACCTCGAACGCATCTACGATTCAGTT	1794
Db	129479	TCCTTACCATTACCAATTCCTGAGTACGATCTACCTCGAACGCATCTACGATTCAGTT	129538
QY	1795	TCTAGTAAAGTACAAAACCGATCCCGCGCAGTGAATCTCTGCTCCAGCCCGTGAAT	1854
Db	129539	TCTAGTAAAGTACAAAACCGATCCCGCGCAGTGAATCTCTGCTCCAGCCCGTGAAT	129598
QY	1855	GGCAAAATCCAAAGCTGAACTCCAGCGTCTGCTGCTTATGGCGCCAGTCTCTGAGGAC	1914
Db	129599	GGCAAAATCCAAAGCTGAACTCCAGCGTCTGCTGCTTATGGCGCCAGTCTCTGAGGAC	129658
QY	1915	TTGACGAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATTTGGTACGATTCAGTCTCC	1974
Db	129659	TTGACGAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATTTGGTACGATTCAGTCTCC	129718
QY	1975	CATCTCCCGGCCAAGATCGGAGATGAGAGGCCACTCCGACCGAGCTTCAAGAACCC	2034
Db	129719	CATCTCCCGGCCAAGATCGGAGATGAGAGGCCACTCCGACCGAGCTTCAAGAACCC	129778
QY	2035	ATGACCTTAAACGGTGTCTAATAGTGCAGACGACGAGTGCACCGAAAGAACGGCTTA	2094
Db	129779	ATGACCTTAAACGGTGTCTAATAGTGCAGACGACGAGTGCACCGAAAGAACGGCTTA	129838
QY	2095	GCGCTCATGTGTCAGTGCAGGCGCTCCGCTGCTCCGCTGCTCCGCTGCTCCGCTGCT	2154
Db	129839	GCGCTCATGTGTCAGTGCAGGCGCTCCGCTGCTCCGCTGCTCCGCTGCTCCGCTGCT	129898
QY	2155	AAGCAACCGTCTCTCTGGAAAGTTGA	2182
Db	129899	AAGCAACCGTCTCTCTGGAAAGTTGA	129926
RESULT 11			
LOCUS	AC146146/c	217022 bp	DNA linear HTG 01-AUG-2003
DEFINITION	Pan troglodytes chromosome UNK clone RP43-2G9, WORKING DRAFT		
SEQUENCE	SEQUENCE, 26 unordered pieces.		
ACCESSION	AC146146		
VERSION	AC146146.1	GI:33387100	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Pan troglodytes (chimpanzee)		
ORGANISM	Pan troglodytes		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.		
AUTHORS	1 (bases 1 to 217022)		
TITLE	Wilson, R.K.		
JOURNAL	The sequence of Pan troglodytes clone		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 217022)		
TITLE	Wilson, R.K.		
JOURNAL	Direct Submission		
COMMENT	Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
----- Project Information -----
Center project name: C_PT002G09
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 204681 bases at least Q40
Consensus quality: 206693 bases at least Q30
Consensus quality: 208289 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1098: contig of 1098 bp in length
* 1099 1198: gap of unknown length
* 1199 2982: contig of 1784 bp in length
* 2983 3082: gap of unknown length
* 3083 5180: contig of 2098 bp in length
* 5181 5280: gap of unknown length
* 5281 7152: contig of 1872 bp in length
* 7153 7252: gap of unknown length
* 7253 9449: contig of 2197 bp in length
* 9450 9549: gap of unknown length
* 9550 10943: contig of 1394 bp in length
* 10944 11043: gap of unknown length
* 11044 13812: contig of 2769 bp in length
* 13813 13913: gap of unknown length
* 13914 13953: contig of 4441 bp in length
* 13954 14854: gap of unknown length
* 14855 21775: contig of 3222 bp in length
* 21776 24307: contig of 2532 bp in length
* 24308 24407: gap of unknown length
* 24408 28212: contig of 3805 bp in length
* 28213 28312: gap of unknown length
* 28313 34694: contig of 6382 bp in length
* 34695 38278: contig of 3484 bp in length
* 38279 38379: gap of unknown length
* 38380 44972: gap of unknown length
* 44973 51522: contig of 6550 bp in length
* 51523 51622: gap of unknown length
* 51623 59330: contig of 7708 bp in length
* 59331 59431: gap of unknown length
* 59432 67686: contig of 8255 bp in length
* 67687 67786: gap of unknown length
* 67787 76304: contig of 8519 bp in length
* 76305 76404: gap of unknown length
* 76405 91685: contig of 15280 bp in length
* 91686 91785: gap of unknown length
* 91786 103782: contig of 11997 bp in length
* 103783 115983: contig of 12102 bp in length
* 115984 116083: gap of unknown length
* 116084 131828: contig of 15745 bp in length
* 131829 131929: gap of unknown length
* 131930 148774: contig of 16846 bp in length
* 148775 168325: gap of unknown length
* 168326 168325: contig of 19451 bp in length

168326 168425: gap of unknown length
168426 185595: contig of 17170 bp in length
185596 185695: gap of unknown length
185696 217022: contig of 31327 bp in length.

FEATURES

source

	source	Location/Qualifiers	length
*	168326	158425: gap of unknown length	
*	168426	185595: contig of 17170 bp in length	
*	185596	185695: gap of unknown length	
*	185696	217022: contig of 31327 bp in length.	
FEATURES			
		1. 217022	
		/organism="Pan troglodytes"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:9598"	
		/chromosome="UNK"	
		/clone="RP43-2G9"	
		1. 1098	
misc_feature		/note="assembly_name:Contig18"	
misc_feature		1199. 2982	
		/note="assembly_name:Contig23"	
misc_feature		3083. 5180	
		/note="assembly_name:Contig25"	
misc_feature		5281. 7152	
		/note="assembly_name:Contig26"	
misc_feature		7253. 9449	
		/note="assembly_name:Contig27"	
misc_feature		9550. 10943	
		/note="assembly_name:Contig28"	
misc_feature		11044. 13812	
		/note="assembly_name:Contig29"	
misc_feature		13913. 18353	
		/note="assembly_name:Contig30"	
misc_feature		18454. 21675	
		/note="assembly_name:Contig31"	
misc_feature		21776. 24307	
		/note="assembly_name:Contig32"	
misc_feature		24408. 28212	
		/note="assembly_name:Contig33"	
misc_feature		28313. 34694	
		/note="assembly_name:Contig34"	
misc_feature		34795. 38278	
		/note="assembly_name:Contig35"	
misc_feature		38379. 44872	
		/note="assembly_name:Contig36"	
misc_feature		44973. 51522	
		/note="assembly_name:Contig37"	
misc_feature		51623. 59330	
		/note="assembly_name:Contig38"	
misc_feature		59431. 67685	
		/note="assembly_name:Contig39"	
misc_feature		67786. 76304	
		/note="assembly_name:Contig40"	
misc_feature		76405. 91684	
		/note="assembly_name:Contig41"	
misc_feature		91785. 103781	
		/note="assembly_name:Contig42"	
misc_feature		103882. 115983	
		/note="assembly_name:Contig43"	
misc_feature		116084. 131828	
		/note="assembly_name:Contig44"	
misc_feature		131529. 148774	
		/note="assembly_name:Contig45"	
misc_feature		148875. 168325	
		/note="assembly_name:Contig46"	
misc_feature		168426. 185595	
		/note="assembly_name:Contig47"	
misc_feature		185596. 217022	
		/note="assembly_name:Contig48"	

ORIGIN

	Query Match	29.6%;	Score 714.4;	DB 2;	Length 217022;
	Best Local Similarity	37.2%;	Pred. No. 3.9e-193;		
	Matches 727;	Conservative 0;	Mismatches 21;	Indels 0;	Gaps 0;
QY	1435	CAGCTTCCTCTC	CATGATGAAGAAAT	CAACTCTCACTTAAATGGGAC	TGGACCACTTGA
Db	150493	CAGGTGCGCTCTG	CTTCTCAGAAATCCACT	CTCTAAATGGGAC	TGGACCACTTGA

RESULT 12	
LOCUS	BC057482
DEFINITION	Danio rerio cDNA clone IMAGE:5915397, partial cds.
ACCESSION	BC057482
VERSION	BC057482.1 GI:34785786
KEYWORDS	Danio rerio (zebrafish)
SOURCE	Danio rerio
ORGANISM	Danio rerio
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio. 1 (bases 1 to 3075) Srausberg,R.L., Reingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zebberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Matusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schetz,T.E., Brownstein,M.J., Usdin,T.B., Teshnyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.W, Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulvik,S.W,
AUTHORS	

/note="COG5077; Region: COG5077, Ubiquitin
 carboxyl-terminal hydrolase [Posttranslational
 modification, protein turnover, chaperones]"
 /db_xref="CDD:COG5077"

ORIGIN	Query Match	25.4%;	Score 613.4;	DB 5;	Length 3075;
	Best Local Similarity	66.1%;	Pred. No. 2.1e-164;	Indels 12;	Gaps 4;
Matches	951;	Conservative	0;	Mismatches 476;	
QY	67	TGAACAATGACCATAGTTGACAAAGCTTCTGAAATCTTTCAGACC---	CATCAGCCTATFCAG	123	
DB	21	TGAATAATGACCATAGTTGACAAACCTCCAGAGAAACCTGACCTGGAGTCAGTCGCGTGC	80		
QY	124	AATCAGCCTGGCAGCTCCGAGGAGCTCTACCTGGAGACATGATGACAGTTCTGCACGC	183		
DB	81	AAGCACTCCAGCTCCATCACACCATTTCTCTCCAGAGACATGGAGACGATAGTTCGAGC	140		
QY	184	TGGGGTGTGTGTCATTGATGATGTCGAAATCAG--ACACTTTCTTTAGGACCA	240		
DB	141	TGGAGTGTGGCCCTTCTGCGACTTGAGGCCCAAGACGCTCTGTATGCGCCCG	200		
QY	241	GTACCTGGTCTGTAGTTTATTCGAGTTCAT---CTGTACTGTGATAAATCAAAACCATCA	297		
DB	201	GCAATTGGAGGGACTGTGTAATGGTAGTAATAACACACGACTCCAAAGTCGAAAGGCCACGA	260		
QY	298	CCACAAAGGATCAAGCCCTAGTGTGGGATCGCTCTCCACAGAAAGTTCTTTCCCA	357		
DB	261	GAGCAAGTGGGATGACCAAGTGGGATGGCATTTCTCTCCCTCAAAAGGTCCTGTTTCC	320		
QY	358	TCTGAGAAGATTTCCTTAAGTGGCAACAACCTCATAGAGTTGAGCTGGGCTCCAGAAAT	417		
DB	321	CCAGAGCGGCTCTGCTTGAAGTGGAAACGAGGCCACGCAATTGGAGCAGGTCTCCATPAC	380		
QY	418	TTGGCCAATACCTGTTTGGCAATGCAAGCACTGCAAGTGTTTAACTACACACCACTCTT	477		
DB	381	CTGGGCAACACATGTTTCCCTGAACTCTACCTCGAGTGTCTGACCTACACTGCTGCCCTT	440		
QY	478	GCCAAATTACATGCCATCACATGAAACACTCCAAAACATGTCATGAGAGGCTTTGTATG	537		
DB	441	GCCAACTACATGCTGACAGAGAGCAATTCCAAACACATGTCATGAGCCCGGTTTTGTATG	500		
QY	538	ATGTGTACATGCAAGCACATATTACCAGGACACTCAGTAAATCCTGGGACGTTATTAAA	597		
DB	501	ATGTGTACCATGCAAAATCAATATCCAAGTGTTCGCACTCTGGAATGTCATTAAAG	560		
QY	598	CCAATGTTTGTATCAATGAGATCGGGGTATAGCTAGGCACCTCCGTTTGTGGAAACCAA	657		
DB	561	CCCATCAGTGTCTTAATGAATTGAAAAGGATGGAAAGCATTTTAGATTTTGGAAAGTCAA	620		
QY	658	GAAGATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAAGACATGCTGAAT	717		
DB	621	GAAGATGCCACGAGTCTTTCAGGTATACACTGTGGATGCTATGCAAAAGTCTTGCCTTCCT	680		
QY	718	GGCAGCATAAATTAGACAGACACACCGAGGCCACCACTCTTGTGTTGTGATATTTTGA	777		
DB	681	G----GAAACAAACTCGACAGGCAAACTCAGGCAACCACTTTGTATCATCAGATTTTGA	737		
QY	778	GGATACCTTAAGATCTAGAGTCAAATGTTTAAATTCGACGGGCTTTCAGATCTTTTGAAT	837		
DB	738	GGATATCTGAGATCAGAGTCAAATGCTGTAATTCGAAAGCAGTCTCTCATACTTTCGAC	797		
QY	838	CCATATCTTGATATAACATTTGAGATAAAGGCTGCTCAGAGTGTCAACGAGCATTTGGAG	897		
DB	798	CCATATTTGGATATTTTCATTTGAGATAAAGACGGCTCAGACACTCTCTAAGGCATTTTGA	857		
QY	898	CAGTTTGTGAAGCCGGMACAGCTTGTATGGAGAAAACCTGCTACAGGTGCAAGTGTGATAA	957		
DB	858	CAATTGTGTAACCTGAGCAACTGATGGGACATGCCCTATAATGCTCCAAATGTAAG	917		
QY	958	AAGATGGTTTCCAGCTTCAAGAGGTTTCACTATCCATAGATCCCTCTAATGTTCTTACATT	1017		
DB	918	AAATGGTTTACTGCTCAAGAGATTTCACCGTTTCATCGCAGTGTCTAATGTTCTCAATC	977		

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

CDS

misc feature


```

QY 1018 TCTCTGAAACGTTTGTGCAAAATTTTACCGGTGGAAATTTCTAGAGATGTGAATACCCCT 1077
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 978 TCACAGAGCGGTTTACCACTTTAATGGAGAAAATAACAAAGGATGTGAGGTACGCA 1037

QY 1078 GAGTATCTTGATATTCGGCCATATATCTCAACCCACGAGAGCCAAATGTCTACGTC 1137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1038 GAGCATCTAGATCTGCGTCCATTCATGTCAGTCTCATGGGAGCCACAAATCTATGCT 1097

QY 1138 TTCTATGCACTGCTGGTCCACACTGTTTTAAATGGCCATGCTGGCCATTAATCTCTGTAC 1197
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1098 CTGTATGCTGTGTGTTCACTTCGCGTTCAGTTGCCACGCTGACATTAATCTATGCTAC 1157

QY 1198 ATAAAGCTAGCAATGGCTCTGATCAATGAATGCACTCCATTTGATCTACAGTGAT 1257
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1158 ATTAAGCTAGCAATGGCGAGTGTATCAGATGAATGACTCGCGGTATCTCTCAGCAGT 1217

QY 1258 ATTAGATCGGTACTACGCAACAGCTATGTCTCTTTTATATCAGGTCCCATGATGTG 1317
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1218 ATAAGAACAGTCTTAACCAACAGCATATCTGCTCTCTACATCAGTCACTGATGTA 1277

QY 1318 AAAAATGGAGGTAACTTACTATCCACCATAGCCCGGCGAGTCTCTCCCGGCC 1377
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1278 AAAAATGGAGGCACTTTAGCCAGATGAACCATCTCCAGGTCACTCTCTCCAGACCA 1337

QY 1378 GTCATCAGTCAGCGGTTGTGACCAACACAGGCTGGCGCAGGCTTTATCGGACACAG 1437
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1338 TCATTTCCACTCAAGATGAATGGGCTCAGTATACACACAGTCTCTCATTTGCCCTCAA 1397

QY 1438 CTTCCCTCTCAATGATAAGAAATCCACCTCACTTAAATGGGACTGGACATTAAGA 1496
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1398 CTGCCACCTCATATGCTGAAGAACTCGTCATATGTTAATGGAAATGTTCTTCAAGGA 1456

```

```

RESULT 13
BC047168
LOCUS
DEFINITION
  Danio rerio, Similar to KIAA1453 protein, clone IMAGE:2601270,
  mRNA, partial cds.
ACCESSION
  BC047168
VERSION
  BC047168.1
KEYWORDS
  GI:28502772
SOURCE
  Danio rerio (zebrafish)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Danio.
  1. (bases 1 to 3274)
  Strausberg,R.
  Direct Submission
  Submitted (21-FEB-2003) National Institutes of Health, Mammalian
  Gene Collection (MGC), Cancer Genomics Office, National Cancer
  Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
  USA
  NIH-MGC Project URL: http://mgc.nci.nih.gov
  Contact: MGC help desk
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
  cDNA Library Preparation: Dr. Sumio Sugano and Dr. Ko-ichi Kawakami
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
  DNA Sequencing by: Institute for Systems Biology
  http://www.systemsbio.org
  contact: amadan@systembiology.org
  Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
  Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
  Clone distribution: MGC clone distribution information can be found
  through the I.M.A.G.E. Consortium/LINL at: http://image.llnl.gov
  Series: IRAP Plate: 101 Row: d Column: 21.
  Location/Qualifiers
    1..3274
    /organism="Danio rerio"
    /mol_type="mRNA"

```

```

/strain="AB"
/db xref="taxon:7955"
/clone="IMAGE:2601270"
/tissue type="Whole body, adult, (one male and one female,
including unfertilized eggs)"
/clone_lib="Sugano Kawakami zebrafish DRA"
/lab_host="DH10B"
/notes="Vector: pME18S-FL3"
225..>3274
/codon_start=1
/product="Similar to KIAA1453 protein"
/protein_id="AAH47168.1"
/db_xref="GI:28502773"
/translation="MESSSSSSVSPSATLSRAKTVCMAPATGATVNGSNNTALOVE
RPEQVMTSGDIALPQKVLFPERLCLKWNOHRIAGLHNLGNTCLFNLSTLOCLT
YTAFLANYMLTREHSTKTCHEPGCMCTWONHIIQVANSNGVIKPISVLNKLKIGK
HPRFGSQDAHELRVTVDAMQKSLCPGNKLDQTOATTFVHQIFGSLKRVKCLNC
KAVSDTFDYLDLSLEIKTAQTLKAFQVPRPEQDGDGNAYKCKKMTVASKRFT
VHRSNVLTSUKRFTNFGKLTQDVYAEHLDLRPFMSQSGEPQIYALYAVLVS
GSCAHGYCYIKASNGQWYQMNDSVLSLDIRTLNQAYLLFIRGPDVKNQGDG
SOMNHTPGGSSPRPSIPLKMGPOYTTTTFIGPOLPMLKNSVYNGSSKEYHGG
SKPNRNFQGVTKIPGSHSTSSASASSSLPVRSMGIPDSKPKLTELIGQKVPRA
QTQSPNCSLSSASHSOSTSSSTSTSGFPQVQVNGHTGASFLVFPQGSSESD
QEAULENGSTYILAPKLTNGVADDAQSDSSLPHTSTKINGSNPHVCENSD
SHVENSQDPTSQNCLSKANGFNHNDKMDIDPHTSSQISDLSCSEMELNLKPAHN
ESRSNPASDDHSCSPSNKRLNLPSESSQSAQIQSPKSLVLTDTDLHTET
ASKVELTORHATHTATGSDLVNNVSLVSHGHTSPHKLKEYVRSSEGESKEAPS
HSDKFGSGDWRSGEDRNRGSDRPHFSTPKDRERVYRHDHSDRSRSGHVSQDSR
RTSREYRDRDLRHWDRFSHRPREHYFORHRDHRDWRDHRFSDSVRPSGTH
NNGVSNHSHRGMEAHGATHVTNGSKRPSVSPHPRHKKRHSFSPVDARESS
DECKAKSKKKRNKDKHRHRELSSEGNSSMRHKKKKKK"

```

ORIGIN

```

Query Match      25.4%; Score 613.4; DB 5; Length 3274;
Best Local Similarity 66.1%; Pred. No. 2.1e-164;
Matches 951; Conservative 0; Mismatches 476; Indels 12; Gaps 4;

QY 67 TGAACAATGACCATAGTTGACAAAAGCTTCTGAATCTTCAGAAC--CATCAGCCTATCAG 123
Db 126 TGAATAATGACCATAGTTGACAAACCTCCAGAGAAACCTGACCTGGAGTCAGTGGGTGC 185

QY 124 ATACAGCTGGCAGCTCCGAGCAGTCTCACTGGAGACATGATGAGTCTGCGACG 183
Db 186 AAGCACCACGCTCCATCACACCATTTCTCTCCAGAGACATGGAGCAGTAGTTCGAGC 245

QY 184 TGGGCTGCTGCTCTTCTTCAATTGATGTGTCAAATCAC--ACACTTCTTTTAGACCA 240
Db 246 TGGAGTGTGGCCCTTCTGCCACTGAGCTTTCAGGGCCAGACGGTCTGTATGGCCCG 305

QY 241 GTACCTGTGTGTAGTTTATTCGAGTTTCAT---CTGTACCTGTATAAATCAAAACATCA 297
Db 306 GCCATTGGAGGACCTGTGAATGGTAGTAATACACAGCACTCCAAGTCGAAGGCCACGA 365

QY 298 CCACAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCA 357
Db 366 GAGCAAGTGGGATGACCAAGTGGGATGGCATGTCTCTCCCTCAAAAGTCTGTGTTTCC 425

QY 358 TCTGAGAAGATTGTCTTAAAGTGGCAAACTCATAGAGTTGGAGCTGGGCTCCAGAT 417
Db 426 CCAGAGCGGCTCTGCTTGAAGTGGNAACCAAGGCCACCCCATTTGGAGCAGGTCTCCATAAC 485

QY 418 TTGGCAATACCTGTTTTGCAATGCGAGTGTGTTTAACTACACACACCACTCTT 477
Db 486 CTGGCAACACATGTTTCTGAACTCTTACCTGAGTGTCTGACCTACACTGCTCCCTT 545

QY 478 GCCAATACATGCTATCATGAACTCCAAACATGTCTCATGAGAGGCTTTGTATG 537
Db 546 GCCAATACATGCTGACGAGAGACATTTCCAAACATGTCTCATGAACCCGGGTTTGTATG 605

QY 538 ATGTGTCAATGCAAGCACATATTACCCAGGCATCTCAGTATCTCTGGGAGCTTTATAA 597
Db 606 ATGTGTACCATGCAAAATACATCATCAAGTGTGTCCAACTCTGGAATGTCTATTAG 665

```

TITLE		Protease homologs	
JOURNAL	Patent: US 6331427-A 223 18-DEC-2001;		
FEATURES	Location/Qualifiers		
source	1..722		
	/organism="unknown"		
ORIGIN	/mol_type="genomic DNA"		
Query Match	19.7%; Score 474.6; DB 6; Length 722;		
Best Local Similarity	92.9%; Pred. No. 1.4e-124;		
Matches 562; Conservative	0; Mismatches 33; Indels 10; Gaps 6;		
QY	1061 AGGATGTGAAATACCTCGTAGTATCTTGATATTGCGGCATATATGCTCAACCCACGGAG 1120		
Db	107 AGGATGTGAAATACCTCGTAGTATCTTGATATTGCGGCATATATGCTCAACCCACGGAG 166		
QY	1121 AGCCAAATGCTCACTGCTTTGTATGCAAGTCTGCTGCACACTGGTTTTAAATTGCCATGCTG 1180		
Db	167 AGCCAAATGCTCACTGCTTTGTATGCAAGTCTGCTGCACACTGGTTTTAAATTGCCATGCTG 226		
QY	1181 GCCATTACTCTCTACATATAAAGCTAGCAATGGCCCTCTGSTATCAAAATGAATGACTCCA 1240		
Db	227 GCCATTACTCTCTACATATAAAGCTAGCAATGGCCCTCTGSTATCAAAATGAATGACTCCA 286		
QY	1241 TTGCTATCTACCACTGATATTAGATCGGTACTCAGCCAAACAGCCTATGTGCTCTTTTATA 1300		
Db	287 TTGCTATCTACCACTGATATTAGATCGGTACTCAGCCAAACAGCCTATGTGCTCTTTTATA 346		
QY	1301 TCAGGTCCTCAATGATGAAAATGGAGGTGAACCTTACTCATCCACCCTAGCCCGGCC 1360		
Db	347 TCAGGTCCTCAATGATGAAAATGGAGGTGAACCTTACTCATCCACCCTAGCCCGGCC 406		
QY	1361 AGTCCTCTCCCGGCGCGGTACATCAGTCAGCGGGTGTGCACCAACAAACAGGCTGCGCCAG 1420		
Db	407 AGTCCTCTCCCGGCGCGGTACATCAGTCAGCGGGTGTGCACCAACAAACAGGCTGCGCCAG 465		
QY	1421 GCCTTTATCGGACCACAGCTTCCCTCTCTCACATGATAAGAAATCCACCTCACTTAAATGGGA 1480		
Db	466 GCCTTTATCGGACCACAGCTTCCCTCTCTCACATGATAAGAAATCCACCTCACTTAAATGGGA 525		
QY	1481 C-TGGACCATTTGAAGACAGCCACAGTTCCTCATGTCGAGTCTTAACGGGAATTCACGT 1539		
Db	526 CATGGACCATTTGAAGACAGCCACAGTTCCTCATGTCGAGTCTTAACGGGAATTCACGT 585		
QY	1540 GTCAACAGGGCTAGTCCCTGTTTAATGCTTCAGCTTCTGTCCAAAACCTGGTCAGTTAATAGG 1599		
Db	586 GTCAACAGGGCTAG-CCTGTTTAATGTT- ---CACTNTGTCCAAAATGGTCAGTTAATAGG 640		
QY	1600 TCCTCAGTGATGCCAGAACATCTCTAAGAACAAAAAATTCAATCAGTATTCACAACAAG 1659		
Db	641 T-CTCAGTGATCC--AGAACATCTTANGAAACAAAAAATTCANAGTATNACACAAGTTG 697		
QY	1660 TTGCC 1664		
Db	698 CTGTC 702		
RESULT 15			
LOCUS	AX867589	521 bp	DNA linear PAT 17-DEC-2003
DEFINITION	Sequence 2494 from Patent EP1074617.		
ACCESSION	AX867589		
VERSION	AX867589.1		
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		
JOURNAL	Primers for synthesising full-length cDNA and their use		
	Patent: EP 1074617-A 2494 07-FEB-2001;		

ACCESSION	AX867589
VERSION	AX867589.1
KEYWORDS	GI:40022452
SOURCE	1 Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ishii, S., Sugiyama, T., Wakamatsu, A., Negai, K. and Otsuki, T.
TITLE	Primers for synthesising full-length cDNA and their use
JOURNAL	Patent: EP 1074617-A 2494 07-FEB-2001;

THIS PAGE BLANK (USPTO)

Copyright (c) 1993 - 2004 CompuGen Ltd.
GENCORE version 3.1.6

Result No.	Score			Match	Length	DB	ID	Description
	Score	Match	Length					
1	2412	100.0	2412	4	AaF81717	Human pro		
2	2371.6	98.3	4659	6	ABK86140	cDNA enco		
3	2365	98.1	3660	4	AAH17633	Human cDN		
4	2339.2	97.0	2347	4	AAH19322	Human ubi		
5	2339.2	97.0	2347	7	ABX93880	Human cDN		
6	2310	95.8	4483	6	AB211508	Human pol		
7	1675.6	69.5	1679	6	AB139772	Human NS		
8	727.2	30.1	1682	4	AAK79798	Human imm		
9	727.2	30.1	1682	4	AAK70010	Human imm		
10	727.2	30.1	21436	4	AAK79799	Human imm		
11	727.2	30.1	21436	4	AAK70011	Human imm		
12	517.4	21.5	1122	4	AAS41077	cDNA enco		
13	493	20.4	577	4	AA541613	cDNA enco		
14	474.6	19.7	722	6	ABK30453	Human G-p		
15	468	19.4	521	4	AAH05659	Human cDN		
16	443	18.4	1593	9	ADC27036	Human deu		
17	436	18.1	1898	9	ADC27022	Human deu		
18	434.6	18.0	1763	7	ACA61698	cDNA enco		
19	433.4	18.0	1593	9	ADC27020	Human deu		
20	433	18.0	4746	4	AAH46504	Human deu		
21	431.8	17.9	1593	9	ADC27014	Human deu		
22	431.2	17.9	3051	9	ADC27060	Human deu		
23	430.2	17.8	1593	9	ADC27028	Human deu		

PT	disorders such as cancer.	Db	601	ATGTTTGTTCATCAATGAGATGCGCGCTATAGCTAGGCACCTCCGCTTTTGGAAACCAAGAA	660
XX		Qy	661	GATGCCCATGAATTCCTTCAATACACTGTGTATGCTATGAGAAAGCATCTTGAATGGC	720
PS	Claim 5; Page 119-120; 134pp; English.	Db	661	GATGCCCATGAATTCCTTCAATACACTGTGTATGCTATGAGAAAGCATCTTGAATGGC	720
XX		Qy	721	AGCAATAAATATAGACAGACACACCCAGGCCACCACTCTTGTTCAGATATTTGGAGGA	780
CC	AAAF81714 to AAF81740 encode the human proteases and protease inhibitors (PPIMs) given in AAB74668 to AAB74694. The PPIMs can have activities such as: anti-human immunodeficiency virus (HIV); antidiabetic; antithyroid; immunostimulant; immunomodulator; antiinflammatory; immunosuppressive; nephrotropic; antitumor; thyromimetic; cytostatic; antibacterial; virucide; antiparasitic; and hepatotropic. PPIM polynucleotide and protein sequences can be used in the diagnosis, treatment and prevention of autoimmune/inflammatory disorders such as AIDS, DiGeorge's syndrome, severe combined immunodeficiency disease (SCID), Chediak-Higashi syndrome, Cushing's disease, Addison's disease, autoimmune thyroiditis, Crohn's disease, diabetes mellitus, Good pasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, Sjogren's syndrome, Werner's syndrome, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and cell proliferative disorder such as arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, psoriasis and cancer. PPIM polynucleotide sequences can be used in somatic or germline gene therapy and in diagnosis of diseases. They can also be used in generating hybridisation probes useful in mapping the naturally occurring genomic sequences and in molecular biology techniques	Db	781	TACTTAAGATCTAGAGTCAAAATGTTAAATGCAAGGCCGCTTTCAGATCTTTGATCCA	840
CC		Qy	841	TATCTTGATATACATTTGGAGATAAGGCTCTCAGAGTGTCAACAGGCAATTCGAGCAG	900
CC		Db	841	TATCTTGATATACATTTGGAGATAAGGCTCTCAGAGTGTCAACAGGCAATTCGAGCAG	900
CC		Qy	901	TTTGTGAAGCCGGAAACAGCTTGTATGGAGAAATCTGTTCAAGTGCACCAAGTGTAAAG	960
CC		Db	901	TTTGTGAAGCCGGAAACAGCTTGTATGGAGAAATCTGTTCAAGTGCACCAAGTGTAAAG	960
CC		Qy	961	ATGTTTCCAGCTTAAAGAGGTTCACTATCCATAGATCCCTTAATGTTCTTACACTTCT	1020
CC		Db	961	ATGTTTCCAGCTTAAAGAGGTTCACTATCCATAGATCCCTTAATGTTCTTACACTTCT	1020
CC		Qy	1021	CTGAAGCGTTTGTCAAAATTTTACCGTGGAAATTTCTAAGGATGTGAATACCTCTGAG	1080
CC		Db	1021	CTGAAGCGTTTGTCAAAATTTTACCGTGGAAATTTCTAAGGATGTGAATACCTCTGAG	1080
CC		Qy	1081	TATCTTGATATTCGCGCATATATGTTCTCAACCCAAACGAGAGCAATTTCTACGCTTG	1140
CC		Db	1081	TATCTTGATATTCGCGCATATATGTTCTCAACCCAAACGAGAGCAATTTCTACGCTTG	1140
CC		Qy	1141	TATCAGTGTCTGCTCAGACTGTTTAAATGCGCATCTGCGCAATCTTCTGCTACATA	1200
CC		Db	1141	TATCAGTGTCTGCTCAGACTGTTTAAATGCGCATCTGCGCAATCTTCTGCTACATA	1200
CC		Qy	1201	AAAGCTAGCAATGTCCTCTGTTATCAATGAATGACATCCATTTATACAGTGTATTT	1260
CC		Db	1201	AAAGCTAGCAATGTCCTCTGTTATCAATGAATGACATCCATTTATACAGTGTATTT	1260
CC		Qy	1261	AGATCGGTACTCA3CCCAACAGCCATGTTGCTCTTTTATATCAAGTCCATGATGTGAA	1320
CC		Db	1261	AGATCGGTACTCA3CCCAACAGCCATGTTGCTCTTTTATATCAAGTCCATGATGTGAA	1320
CC		Qy	1321	AATGAGGTGAACTTACTCATCCACCATAGCCCGCCAGTCTCTCCCGCCCGCTC	1380
CC		Db	1321	AATGAGGTGAACTTACTCATCCACCATAGCCCGCCAGTCTCTCCCGCCCGCTC	1380
CC		Qy	1381	ATCAGTCAAGCGGTGTTGTCAACCAACAGGCTCGCCAGGCTTTATCGGACCAAGCTT	1440
CC		Db	1381	ATCAGTCAAGCGGTGTTGTCAACCAACAGGCTCGCCAGGCTTTATCGGACCAAGCTT	1440
CC		Qy	1441	CCCTCTCATATGATAAAGAATCCACCTCACTTAAATGGGACTGGACCAATTTGAAGACAG	1500
CC		Db	1441	CCCTCTCATATGATAAAGAATCCACCTCACTTAAATGGGACTGGACCAATTTGAAGACAG	1500
CC		Qy	1501	CAAGCAGTTCATGTCAGGTCCTTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGT	1560
CC		Db	1501	CAAGCAGTTCATGTCAGGTCCTTAAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGT	1560
CC		Qy	1561	AATGCTTCAGCTTCTGTCCAAATTCAGTGTCAATAGGTCCTCAGTGTATCCAGAAAT	1620
CC		Db	1561	AATGCTTCAGCTTCTGTCCAAATTCAGTGTCAATAGGTCCTCAGTGTATCCAGAAAT	1620
CC		Qy	1621	CCTAAGAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1680
CC		Db	1621	CCTAAGAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1680
CC		Qy	1681	TCTCAACCTAACCTTTCATAGTAATTTCTTTGGAGAAACCTTACCAAGCCCTTCTCT	1740
CC		Db	1681	TCTCAACCTAACCTTTCATAGTAATTTCTTTGGAGAAACCTTACCAAGCCCTTCTCT	1740

XX		Qy	1	GGGCTGTGTGGCGGGCGGGCGGGCGGGCGGGGATGGAGCGCGCGAGCGGGT	60
PS		Db	1	GGGCTGTGTGGCGGGCGGGCGGGCGGGCGGGGATGGAGCGCGCGAGCGGGT	60
XX		Qy	61	CAGAGTTGAAACATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT	120
PS		Db	61	CAGAGTTGAAACATGACCATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT	120
XX		Qy	121	CAGAAATCAGCTGCGAGCTCCGAGCGAGTCTACCTGGAGACATGGATGAGGTTCTGCC	180
PS		Db	121	CAGAAATCAGCTGCGAGCTCCGAGCGAGTCTACCTGGAGACATGGATGAGGTTCTGCC	180
XX		Qy	181	AGCTGGGGTGTGTCTTCAATGAAATGATGTGCAAAATGATGTTTCTTTAGGACCA	240
PS		Db	181	AGCTGGGGTGTGTCTTCAATGAAATGATGTGCAAAATGATGTTTCTTTAGGACCA	240
XX		Qy	241	GTACTGCTGTGTAGTTTATTCGAGTTTCACTGTACCTGATATAATCAAAACCATCACCA	300
PS		Db	241	GTACTGCTGTGTAGTTTATTCGAGTTTCACTGTACCTGATATAATCAAAACCATCACCA	300
XX		Qy	301	CAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTTCCCATCT	360
PS		Db	301	CAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTTCCCATCT	360
XX		Qy	361	GAGAGATTTCTTAACTGGGCAACAACTCATAGTTGGAGTGGCTCCAGAAATTG	420
PS		Db	361	GAGAGATTTCTTAACTGGGCAACAACTCATAGTTGGAGTGGCTCCAGAAATTG	420
XX		Qy	421	GGCAATACCTGTTTGGCAATGACAGTGTGAGTGTGTTAACTCAGACACCACTCTGCC	480
PS		Db	421	GGCAATACCTGTTTGGCAATGACAGTGTGAGTGTGTTAACTCAGACACCACTCTGCC	480
XX		Qy	481	AATATACGCTATCATGATGAACATCTCAAAACATGTCATGAGAGGCTTTTGTATGATG	540
PS		Db	481	AATATACGCTATCATGATGAACATCTCAAAACATGTCATGAGAGGCTTTTGTATGATG	540
XX		Qy	541	TGTAAATGAGGACATATTAACCGGCACTCAGTAATCTCTGGGCACTTATTAACCA	600
PS		Db	541	TGTAAATGAGGACATATTAACCGGCACTCAGTAATCTCTGGGCACTTATTAACCA	600
XX		Qy	601	ATGTTTGTTCATCAATGAGATGGGGTATPAGTCTAGGCACTCTCGTTTGGAAACCAAGAA	660

Query Match 100.0%; Score 2412; DB 4; Length 2412;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2	Key	Location/Qualifiers
ABK86140		
ID	ABK86140	standard; cDNA; 4659 BP.
XX	ABK86140;	
XX	AC	
XX	DT	26-AUG-2002 (first entry)
XX	XX	cDNA encoding human protease PRPS7.
XX	DE	
XX	DE	
KW	KW	Protease; human; gastrointestinal disorder; gastritis; atherosclerosis;
KW	KW	ulcerative colitis; Reye's syndrome; cardiovascular disorder; gene; ss;
KW	KW	hypertension; myocardial infarction; autoimmune disease; AIDS; PRPS;
KW	KW	inflammatory disorder; acquired immunodeficiency syndrome; asthma;
KW	KW	Grave's disease; cell proliferative disorder; hepatitis; psoriasis;
KW	KW	leukaemia; developmental disorder; Cushing's syndrome; impotence;
KW	KW	epithelial disorder; dermatitis; scabies; eczema; neurological disorder;
KW	KW	Parkinson's disease; dementia; Alzheimer's disease; infertility;
KW	KW	Huntington's disease; multiple sclerosis; reproductive disorder.
OS	OS	Homo sapiens.
XX	XX	
PH	PH	

QY	60
Db	142
QY	120
Db	202

Db	1021	ACGTTTTCGAAAATTTTACCGGTGGAAAAATTCGTAAGGATGTGAAATATACCTCGAGTATCT	1080
Qy	1086	TGATATTTCGGCCATATATGTCTCAACCCAAACGAGAGCAAATTTGTCTAGCTCTGTATGC	1145
Db	1081	TGATATTTCGGCCATATATGTCTCAACCCAAACGAGAGCAAATTTGTCTAGCTCTGTATGC	1140
Qy	1146	AGTCTGGTCCACACACTGGTTTTAAATTGGCAATGCTGGCCAAATTCTTCTGTACATAAAGC	1205
Db	1141	AGTCTGGTCCACACACTGGTTTTAAATTGGCAATGCTGGCCAAATTCTTCTGTACATAAAGC	1200
Qy	1206	TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTTGTATCTTACCAAGTGATATTAGATC	1265
Db	1201	TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTTGTATCTTACCAAGTGATATTAGATC	1260
Qy	1266	GGTACTCAGCCAAACAAGCCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAATGG	1325
Db	1261	GGTACTCAGCCAAACAAGCCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAATGG	1320
Qy	1326	AGGTGAACCTTACTCATCCCAACCCATAGCCCGGCCAGTCTCTCTCCGCGCCGTCTATCAG	1385
Db	1321	AGGTGAACCTTACTCATCCCAACCCATAGCCCGGCCAGTCTCTCTCCGCGCCGTCTATCAG	1380
Qy	1386	TCAGCGGGTGTACCAACAAACAGCTGCGCCAGAGCTTTATCGAGACACAGCTTCCCTC	1445
Db	1381	TCAGCGGGTGTACCAACAAACAGCTGCGCCAGAGCTTTTATCGAGACACAGCTTCCCTC	1440
Qy	1446	TCATGATTAAGAAATCCACCTCACTTAAATGGGACTGGACCAATTGAAAGACACGCCAAG	1505
Db	1441	TCATGATTAAGAAATCCACCTCACTTAAATGGGACTGGACCAATTGAAAGACACGCCAAG	1500
Qy	1506	CAGTTCCATGTCGAGTCCTAAACGGGAATCCAGTGTCAAACAGGGCTAGTCTCTTTAATGC	1565
Db	1501	CAGTTCCATGTCGAGTCCTAAACGGGAATCCAGTGTCAAACAGGGCTAGTCTCTTTAATGC	1560
Qy	1566	TTGAGCTTCTGTCCAAAACCTGGTCAGTTAATAGTCTCTCAGTGATCCCAAGACATCCTAA	1625
Db	1561	TTGAGCTTCTGTCCAAAACCTGGTCAGTTAATAGTCTCTCAGTGATCCCAAGACATCCTAA	1620
Qy	1626	GAAACAAAATATCAATCAGTATTTCACAAAGGTTGCCCTGTTCCGACAGTGTCAAGTCTCA	1685
Db	1621	GAAACAAAATATCAATCAGTATTTCACAAAGGTTGCCCTGTTCCGACAGTGTCAAGTCTCA	1680
Qy	1686	ACCTAACCTTTCATAGTAAATCTTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCAT	1745
Db	1681	ACCTAACCTTTCATAGTAAATCTTTTGGAGAACCTTACCAAGCCGTTCCCTCTCTACCAT	1740
Qy	1746	TACCAATTCTGAGTACAGTCTACTCGAACGATCTTACGATGTCAAGTCTCTAGTAAAGT	1805
Db	1741	TACCAATTCTGAGTACAGTCTACTCGAACGATCTTACGATGTCAAGTCTCTAGTAAAGT	1800
Qy	1806	RACAAAACCGATCCCCCGCAGTGAATCTGCTCCAGCCGCTGATGAATGGCAAAATCCAA	1865
Db	1801	RACAAAACCGATCCCCCGCAGTGAATCTGCTCCAGCCGCTGATGAATGGCAAAATCCAA	1860
Qy	1866	GCTGAACCTCCAGCGTCTGGTGCCCTATGGCGCCGAGTCTCTCTCAGGACCTGTACGAGGA	1925
Db	1861	GCTGAACCTCCAGCGTCTGGTGCCCTATGGCGCCGAGTCTCTCTCAGGACCTGTACGAGGA	1920
Qy	1926	GTCAAAGGGCTGGGCAAGGAGAAATGGGATTTGATACGATTTGTGAGTCTCCACTCTCCCG	1985
Db	1921	GTCAAAGGGCTGGGCAAGGAGAAATGGGATTTGATACGATTTGTGAGTCTCCACTCTCCCG	1980
Qy	1986	CCAAAGTCGGAAGATGAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCTTAAA	2045
Db	1981	CCAAAGTCGGAAGATGAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCTTAAA	2040
Qy	2046	CGGTGCTAATAGTGCAGACAGCAGTAGTGAOCCGAAAGAAACGGCCTTAGCGCTGTATGG	2105
Db	2041	CGGTGCTAATAGTGCAGACAGCAGTAGTGAOCCGAAAGAAACGGCCTTAGCGCTGTATGG	2100
Qy	2106	TGCGAGCTGCCAAGGCCAGCCTGGCCTGCACCTACAGAAATCCCTTTGCTAAGGCAAAACGG	2165
Db	2101	TGCGAGCTGCCAAGGCCAGCCTGGCCTGCACCTACAGAAATCCCTTTGCTTTGGCAAAACGG	2160

Qy	2166	TCTTCTCGAAAGTTCATGCTGCTCTCTTTGCTGTCTCTCCGAGAAGACAAAATCTTTAGA	2222
Db	2161	TCTTCTCGAAAGTTCATGCTGCTCTCTTTGCTGTCTCTCCGAGAAGACAAAATCTTTAGA	2220
Qy	2226	GACCTTCAGGCTTTAGCAACAACTGAAAGGCTCGACGGATGAAATGAGTGCACCTGGAGC	2285
Db	2221	GACCTTCAGGCTTTAGCAACAACTGAAAGGCTCGACGGATGAAATGAGTGCACCTGGAGC	2280
Qy	2286	AGAGAGGGGCCCTCCGAGGACCGCGACCGGAGCTTCAGCTTGGCGAGCCCCCGCGCGA	2345
Db	2281	AGAGAGGGGCCCTCCGAGGACCGCGACCGGAGCTTCAGCTTGGCGAGCCCCCGCGCGA	2340
Qy	2346	ATCCCTGGAGAGCCAGATCGGCGCGCAAGCTT	2378
Db	2341	ATCCCTGGAGAGCCAGATCGGCGCGCGCT	2373
RESULT 4			
AAH19322			
ID	AAH19322 standard; cDNA; 2347 BP.		
XX	AC	AAH19322;	
XX	DT	20-JUL-2001 (first entry)	
XX	DE	Human ubiquitin protease 23431 coding sequence.	
XX	KW	Human; ubiquitin protease; deubiquitinating enzyme; tuberculostatic;	
XX	KW	antiasthmatic; antiinflammatory; antidiarrhoeic; hepatotropic;	
XX	KW	gynaecological; cystostatic; antimicrobial; neuroprotective; anti-HIV;	
XX	KW	immunosuppressive; cardiast; antianaemic; nephrotropic; antibacterial;	
XX	OS	anti-thyroid; gastro-intestinal; ss.	
XX	OS	Homo sapiens.	
XX	FH	Key	
XX	FT	CDS	
XX	FT	Location/Qualifiers	
XX	FT	61..2346	
XX	FT	/*tag= a	
XX	FT	/partial	
XX	FT	/product= "Human ubiquitin protease"	
XX	FT	/note= "No stop codon given"	
XX	PN	WO200123589-A2.	
XX	XX	05-APR-2001.	
XX	PD	29-SEP-2000; 2000WO-US026962.	
XX	PF	29-SEP-1999; 99US-00407356.	
XX	PR	(MILL-) MILLENNIUM PHARM INC.	
XX	PA	Kapeller-Libermann R;	
XX	PI	WPI; 2001-374253/39.	
XX	DR	P-PSDB; AAB82177.	
XX	DR	New human ubiquitin protease, a member of the mammalian deubiquitinating	
XX	PT	enzymes is useful for diagnosis and treatment of e.g. tuberculosis and	
XX	PT	Alzheimer's disease.	
XX	XX	Claim 1; Fig 1; 11lpp; English.	
XX	CC	The present sequence is the coding sequence for human ubiquitin protease	
XX	CC	23431. The ubiquitin protease coding sequence and protein are useful as	
XX	CC	reagents or targets in assays for treatment and diagnosis of ubiquitin-	
XX	CC	mediated or -related disorders, especially disorders mediated by	
XX	CC	deubiquitinating enzymes. The protein and coding sequence are also useful	
XX	CC	for treating disorders involving the following: the spleen e.g.	
XX	CC	tuberculosis and typhoid fever, the lung such as bronchial asthma, the	
XX	CC	colon e.g. as diarrhoea and dysentery, the liver e.g. jaundice and	
XX	CC	cholestasis, the uterus and endometrium e.g. endometritis the brain e.g.	

CC acute meningitis and Alzheimer disease, T-cells including transplant
CC rejection and autoimmune diseases such as systemic lupus erythematosus,
CC diseases of the skin such as malignant melanoma, the bone marrow e.g. B-
CC and T-lymphoid leukemias, the heart including myocardial infarction, red
CC cells e.g. anaemia, the thymus e.g. Hodgkin disease, B-cells e.g.
CC peripheral B-cell neoplasms, the kidney e.g. polycystic kidney disease,
CC the breast including periductal mastitis, the testis and epididymis e.g.
CC syphilis, the prostate e.g. nodular hyperplasia, the thyroid, e.g.
CC hyperthyroidism, the skeletal muscle e.g. rhabdomyosarcoma, the pancreas
CC e.g. ectopic pancreas, reduced platelet number e.g. HIV (human
CC immunodeficiency virus)-associated thrombocytopaenia and disorders
CC involving precursor T-cell neoplasms including precursor T lymphoblastic
CC leukaemia/lymphoma
XX
SQ Sequence 2347 BP; 647 A; 606 C; 528 G; 566 T; 0 U; 0 Other;
Query Match 97.08; Score 2339.2; DB 4; Length 2347;
Best Local Similarity 99.94; Pred. No. 0;
Matches 2341; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 16 GCGGCGGCGCGCGGAGGGGAGTGGAGCGAGCGCGGAGCGGGTCAAGATGCAACAATG 75
DB 4 GCGTCCGCGCGCGCGAGGGGAGTGGAGCGAGCGCGGAGCGGGTCAAGATGCAACAATG 63
QY 76 ACCATAGTTGACAAAGCTTCTGAACTCTTCAGAACCCATCAGCCCTATCAGAAATCAGCCTGGC 135
DB 64 ACCATAGTTGACAAAGCTTCTGAACTCTTCAGAACCCATCAGCCCTATCAGAAATCAGCCTGGC 123
QY 136 AGCTCGGAGGAGTCTACCTGGAGACATGGATGACAGGTTTCGACGCTGGGTCGTGTG 195
DB 124 AGCTCGGAGGAGTCTACCTGGAGACATGGATGACAGGTTTCGACGCTGGGTCGTGTG 183
QY 196 TCTTCAATCAATGATGTGTCAAAATCACACACTTTCTTTAGGACCAGTACTGCTGTGTA 255
DB 184 TCTTCAATCAATGATGTGTCAAAATCACACACTTTCTTTAGGACCAGTACTGCTGTGTA 243
QY 256 GTTTATTCGAGTTCATCTGTACTGTAAATCAAAACCATCACCAAAAGGATCAAGCC 315
DB 244 GTTTATTCGAGTTCATCTGTACTGTAAATCAAAACCATCACCAAAAGGATCAAGCC 303
QY 316 CTAGTGATGGATGCTCTCTCCACAGAAAGTTCTTTCCCATCTGAGAAAGTTTGTCTT 375
DB 304 CTAGTGATGGATGCTCTCTCCACAGAAAGTTCTTTCCCATCTGAGAAAGTTTGTCTT 363
QY 376 AAGTGGCAACAAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTT 435
DB 364 AAGTGGCAACAAACTCATAGAGTTGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTT 423
QY 436 GCCAATGACGACTGACAGTGTAACTACACACCACTCTTGGCCAAATACATGCTATCA 495
DB 424 GCCAATGACGACTGACAGTGTAACTACACACCACTCTTGGCCAAATACATGCTATCA 483
QY 496 CATGACACTCCAAACATGTCATGACAGAGGCTTTGATGATGTACAAATGCAAGCA 555
DB 484 CATGACACTCCAAACATGTCATGACAGAGGCTTTGATGATGTACAAATGCAAGCA 543
QY 556 CATATTACCCAGGCACTCAGTAATCCTCGGGAGTTTAAACCAATGTTGTATCAAT 615
DB 544 CATATTACCCAGGCACTCAGTAATCCTCGGGAGTTTAAACCAATGTTGTATCAAT 603
QY 616 GAGATGCGGCGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCCATGAATTC 675
DB 604 GAGATGCGGCGTATAGTAGGCACTCCGTTTGGAAACCAAGAGATGCCCATGAATTC 663
QY 676 CTTCAATACACTGTTGATGCTATGAGAAAGCATGCTTTGAATGGCAGCAATAAATAGAC 735
DB 664 CTTCAATACACTGTTGATGCTATGAGAAAGCATGCTTTGAATGGCAGCAATAAATAGAC 723
QY 736 AGACACACCCAGGCCACCACTCTGTTGTAGATTTTGGAGGATACCTAAGATCTAGA 795
DB 724 AGACACACCCAGGCCACCACTCTGTTGTAGATTTTGGAGGATACCTAAGATCTAGA 783
QY 796 GTCAAAATGTTTAAATTTGCAAGGGCGTTTTCAGATACCTTTTGTATCCATATCTTTGATATAACA 855

DB 784 GTCAAAATGTTTAAATTTCAAGGGCGTTTCAGATACCTTTTGATCCATATCTTGATATAACA 843
QY 856 TTGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCAATTTGGAGCAGTTTGTGAAGCCGAA 915
DB 844 TTGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCAATTTGGAGCAGTTTGTGAAGCCGAA 903
QY 916 CAGCTTATGAGGAGAAACTCTGTAACAAGTGCAGCAAGTGTAAAAGATGTTTCAGACTTCA 975
DB 904 CAGCTTATGAGGAGAAACTCTGTAACAAGTGCAGCAAGTGTAAAAGATGTTTCAGACTTCA 963
QY 976 AAGAGTTCACATATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAAAGCTTTTGA 1035
DB 964 AAGAGTTCACATATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAAAGCTTTTGA 1023
QY 1036 AATTTTACCGGTGGAATAATTTGCTAAGGATGTCAAAATACCTCAGTATCTTGATATTCGG 1095
DB 1024 AATTTTACCGGTGGAATAATTTGCTAAGGATGTCAAAATACCTCAGTATCTTGATATTCGG 1083
QY 1096 CMTATATGTTCTCAACCCCAACCGAGAGCCAAATGTTCTAGCTTCTTGTATGCAGTGTGTC 1155
DB 1084 CCATATATGTTCTCAACCCCAACCGAGAGCCAAATGTTCTAGCTTCTTGTATGCAGTGTGTC 1143
QY 1156 CACACTGGTTTTAAATTTGCCATGCTGCCAATTACTTCTGTACATAAAAGCTAGCAATGGC 1215
DB 1144 CACACTGGTTTTAAATTTGCCATGCTGCCAATTACTTCTGTACATAAAAGCTAGCAATGGC 1203
QY 1216 CTCTGTATCAAAATGAATGACTTCCATTGTTATCTACAGTGATATTAGATCGGTACTCAGC 1275
DB 1204 CTCTGTATCAAAATGAATGACTTCCATTGTTATCTACAGTGATATTAGATCGGTACTCAGC 1263
QY 1276 CAACAAGCCTATGTCTCTTTTATATCAGGTCCTCATGATGTGAAAATGGAGTGAACCTT 1335
DB 1264 CAACAAGCCTATGTCTCTTTTATATCAGGTCCTCATGATGTGAAAATGGAGTGAACCTT 1323
QY 1336 ACTCATCCCAACCATAGCCCCGGCAGTCTCTTCCCGCCCGCTCATCATGTCAGCGGTT 1395
DB 1324 ACTCATCCCAACCATAGCCCCGGCAGTCTCTTCCCGCCCGCTCATCATGTCAGCGGTT 1383
QY 1396 GTCCACCAACAAACAGGCTGCGCAGGCTTTATCGGACACACAGCTCCCTCTCACATGATA 1455
DB 1384 GTCCACCAACAAACAGGCTGCGCAGGCTTTATCGGACACACAGCTCCCTCTCACATGATA 1443
QY 1456 AAGAATCCACTCACTTAAATGGGACTGGACCAATTTGAAAGACACGCCAAGCAGTCCATG 1515
DB 1444 AAGAATCCACTCACTTAAATGGGACTGGACCAATTTGAAAGACACGCCAAGCAGTCCATG 1503
QY 1516 TCGAGTCTTAAGCGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGGTTCT 1575
DB 1504 TCGAGTCTTAAGCGGAATTTCCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGGTTCT 1563
QY 1576 GTCCAAAACCTGCTCAGTTTAAATAGGTCCTCAGTATGCCAGAACATCCCTAAGAAAACAAA 1635
DB 1564 GTCCAAAACCTGCTCAGTTTAAATAGTCTCTCAGTATCCAGAACATCCCTAAGAAAACAAA 1623
QY 1636 ATTCAATCAGTATTTCAACAACAAGTGTCTGTTCGCCAGTGTCTCAGTCTCAACCTAACCTT 1695
DB 1624 ATTCAATCAGTATTTCAACAACAAGTGTCTGTTCGCCAGTGTCTCAGTCTCAACCTAACCTT 1683
QY 1696 CATAGTAATCTTTGGAGAACCTTACCAGCGGTTCCCTCTTCTTACCATTTACCAATTTCT 1755
DB 1684 CATAGTAATCTTTGGAGAACCTTACCAGCGGTTCCCTCTTCTTACCATTTACCAATTTCT 1743
QY 1756 GCAGTACAGTCTACTCTGAAGCATCTACGATGTCTAGTTCCTAGTAAAGTAAACAAACCG 1815
DB 1744 GCAGTACAGTCTACTCTGAAGCATCTACGATGTCTAGTTCCTAGTAAAGTAAACAAACCG 1803
QY 1816 ATCCCCCGCAGTGAATCCTGCTCCAGCCCCGTGATGAATGGCAAAATCCAAAGCTGAACCTCC 1875
DB 1804 ATCCCCCGCAGTGAATCCTGCTCCAGCCCCGTGATGAATGGCAAAATCCAAAGCTGAACCTCC 1863
QY 1876 AGCGTCTGTGTCCTTATGGCGCGGCTTCTCTGAGGACTCTGACGAGGAGTCAAGGGG 1935

Db 1864 ACGGTCGCTGGTCCCTATGGCGCGAGTCTCTGAGACTCTGACGAGGATCAAGGGG 1923
QY 1936 CTGGCAAGGAGAAATGGGATTTGATGATGAGCTCCACTCTCCCGCCAAAGATGCC 1995
Db 1924 CTGGCAAGGAGAAATGGGATTTGATGATGAGCTCCACTCTCCCGCCAAAGATGCC 1983
QY 1996 GAAGATGAGAGGCCACTCTGGCAGAGCTTCAAGAACCCATGACCTTAACCGTGTAAAT 2055
Db 1984 GAAGATGAGAGGCCACTCTGGCAGAGCTTCAAGAACCCATGACCTTAACCGTGTAAAT 2043
QY 2056 AGTGCACAGCAGCAGTACCGCAAGAAAGCGCTAGCGCTGATGTGTCAGCTGC 2115
Db 2044 AGTGCACAGCAGCAGTACCGCAAGAAAGCGCTAGCGCTGATGTGTCAGCTGC 2103
QY 2116 CAAGGCGAGCGCTCCCTGCACTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCTGGA 2175
Db 2104 CAAGGCGAGCGCTCCCTGCACTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCTGGA 2163
QY 2176 AAGTTGATGCTGCTCTTTGCTGCTCTCCAGAGCAAAATCTTAGAGACTTCCAGG 2235
Db 2164 AAGTTGATGCTGCTCTTTGCTGCTCTCCAGAGCAAAATCTTAGAGACTTCCAGG 2223
QY 2236 CTTAGCAACAACTGAAGGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGC 2295
Db 2224 CTTAGCAACAACTGAAGGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGC 2283
QY 2296 CTTCCGAGGACCGCAGCGAGCTCAGCTGCGAGCGCCCGCCGCGAATCCTCGAG 2355
Db 2284 CTTCCGAGGACCGCAGCGAGCTCAGCTGCGAGCGCCCGCCGCGAATCCTCGAG 2343
QY 2356 GAGC 2359
Db 2344 GAGC 2347

RESULT 5

ABX93880
ID ABX93880 standard; cDNA; 2347 BP.
AC ABX93880;
XX
XX 10-JUN-2003 (first entry)
DE Human cDNA encoding ubiquitin protease.
XX Human; ss; gene; ubiquitin protease; spleen disorder; Hodgkin's disease;
KW lung disorder; adult respiratory distress syndrome; colon disorder;
KW inflammatory bowel disease; liver disorder; jaundice; uterine disorder;
KW endometriosis; brain disorder; Alzheimer's disease; thyroid disorder;
KW acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria;
KW heart disorder; ischaemic heart disease; blood vessel disorder;
KW atherosclerosis; red blood cell disorder; anaemia; thymus disorder;
KW polycystic kidney disease; glomerulonephritis; kidney disorder;
KW testicular disorder; sexually transmitted disease; breast disorder; mastitis;
KW hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder;
KW whipple disease; tumour; cancer.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 61..2347
FT /*tag= a
FT /product= "Ubiquitin protease"
FT /partial
FT /note= "No stop codon shown"
XX
XX US2003037350-A1.
XX
XX 20-FEB-2003.
XX
XX 05-JUN-2002; 2002US-00163547.
XX

29-SEP-1999; 99US-00407356.
05-NOV-1999; 99US-00435311.
01-FEB-2000; 2000US-00496005.
11-FEB-2000; 2000US-0182009P.
14-FEB-2000; 2000US-0182408P.
28-FEB-2000; 2000US-0185503P.
02-NOV-2000; 2000US-00704918.
12-FEB-2001; 2001US-00781598.
14-FEB-2001; 2001US-00782952.
28-FEB-2001; 2001US-00796100.
(MILL-) MILLENNIUM PHARM INC.
Glucksmann MA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;
WPI; 2003-342047/32.
P-PSDB; ABU08951.
New nucleic acid or polypeptide, useful for preparing a composition for
diagnosing or treating diseases e.g., tumor.
Claim 2; Fig 1; 281pp; English.

The invention relates to a new isolated nucleic acid molecule encoding
one of 6 polypeptides (ubiquitin protease, lipase, dynamin, short chain
dehydrogenase, ADAM-TS (a disintegrin and metalloprotease domain protein
with thrombospondin (TS) domains) and gamma butyrobetaine-hydroxylase
(gamma BH). Also included are the polypeptide, host cells containing the
nucleic acid, an antibody that selectively binds to the polypeptide, a
method for producing the polypeptide, a method for detecting the presence
of the polypeptide or the nucleic acid in a sample, a method for
identifying a compound that binds to the polypeptide. The nucleic acid
modulating the activity of the polypeptide and a method for diagnosing or
or polypeptide is useful for preparing a composition for diagnosing or
treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's
disease), lung disorders (e.g. adult respiratory distress syndrome,
pulmonary oedema, chronic bronchitis and emphysema), colon disorders
(e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),
liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver
disease), uterine and endometrial disorders (e.g. endometriosis and
menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's
disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell
disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders
(e.g. urticaria, dermatitis and lupus erythematosus), heart disorders
(e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),
blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and
Raynaud disease), red blood cell disorders (e.g. anaemia), thymus
disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),
kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),
breast disorders (e.g. mastitis), testicular disorders (e.g. sexually
transmitted diseases and cryptorchidism), thyroid disorders (e.g.
hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal
disorders (e.g. whipple disease), as well as tumours and cancers of the
above listed organs/cells. Many more diseases and disorders are listed in
the specification. The present sequence encodes human ubiquitin protease

Sequence 2347 BP; 647 A; 606 C; 528 G; 566 T; 0 U; 0 Other;
Query Match 57.0%; Score 2339.2; DB 7; Length 2347;
Best Local Similarity 59.9%; Pred. No. 0;
Matches 2341; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 16 GCGGCGGCGCGCGCGGATGAGCGAGCGCGCGCGGTCAGAGTTGAACAATG 75
Db 4 GCGTCCGCGCGCGCGCGGATGAGCGAGCGCGCGCGGTCAGAGTTGAACAATG 63
QY 76 ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCATCAGCTATCAGATCAGCTGCG 135
Db 64 ACCATAGTTGACAAAGCTTCTGAATCTTCAGACCATCAGCTATCAGATCAGCTGCG 123
QY 136 AGCTCGAGGCAATCTCACCTGGAGCATGATGAGGTTCTGCCAGCTGGGTGCTGTG 195

Db 124 AGCTCCGAGCAGTCTCACCTGGAGACATGGATGCAGGTTCTGCCAGCTGGGGTGCTGTG 183
QY 196 TCTTCAATCGAATCATGTGCAAAATCACACACATTTCTTTAGGACCACTACCTGGTGCTGTA 255
Db 184 TCTTCAATCGAATCATGTGCAAAATCACACACATTTCTTTAGGACCACTACCTGGTGCTGTA 243
QY 256 GTTATTTCCAGTTCATCTGTACCTCTGATAAATAAATAAACCACATCACCAAAAAGGATCAAGCC 315
Db 244 GTTATTTCCAGTTCATCTGTACCTCTGATAAATAAATAAACCACATCACCAAAAAGGATCAAGCC 303
QY 316 CTAGGTGATGGCANTGCTCTCTCCACAGAAAGTTCCTTTCCCATCTGAGAGAAATTTGCTCTT 375
Db 304 CTAGGTGATGGCANTGCTCTCTCCACAGAAAGTTCCTTTCCCATCTGAGAGAAATTTGCTCTT 363
QY 376 AAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTCTT 423
Db 364 AAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTCTT 423
QY 436 GCCAATGCAAGCACTGCAGTGTGTTAACTTACACACACACACCTCTTGGCCAAATTACATGCTATCA 495
Db 424 GCCAATGCAAGCACTGCAGTGTGTTAACTTACACACACACACCTCTTGGCCAAATTACATGCTATCA 483
QY 496 CATGAACACTCCAAACATGTCAATGCAAGAGCTTTTGTATGATGTGTACAATGCAAGCA 555
Db 484 CATGAACACTCCAAACATGTCAATGCAAGAGCTTTTGTATGATGTGTACAATGCAAGCA 543
QY 556 CATATTTACCCAGGCACTCAGTAATCCTGGGGAGGTTATTTAAACCAATGTTTGTCTATCAAT 615
Db 544 CATATTTACCCAGGCACTCAGTAATCCTGGGGAGGTTATTTAAACCAATGTTTGTCTATCAAT 603
QY 616 GAGATGCGGCGTATAGTAGGCACTTCGTTTGGAAACCAAGAGATGCCCATGAATTC 675
Db 604 GAGATGCGGCGTATAGTAGGCACTTCGTTTGGAAACCAAGAGATGCCCATGAATTC 663
QY 676 CTTCAATACACTGTTGATGTATGAGAAAGCATGCTTGAATGGCAGCAATAAATAGAC 735
Db 664 CTTCAATACACTGTTGATGTATGAGAAAGCATGCTTGAATGGCAGCAATAAATAGAC 723
QY 736 AGACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGA 795
Db 724 AGACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGA 783
QY 796 GTCAAAATGTTTAAATTTGAAGGGCGTTTCAGATACCTTTTGATCCATATCTTTGATATAACA 855
Db 784 GTCAAAATGTTTAAATTTGAAGGGCGTTTCAGATACCTTTTGATCCATATCTTTGATATAACA 843
QY 856 TTGAGATAAAGGCTGTCTCAGAGTGTCAACAGGCAATGGAGCAGTTCGTGAAGCCGGAA 915
Db 844 TTGAGATAAAGGCTGTCTCAGAGTGTCAACAGGCAATGGAGCAGTTCGTGAAGCCGGAA 903
QY 916 CAGCTTGATGAGAAAACTCTGTACAGTGCAGCAGTGTAAAGATGGTTCCAGCTTCA 975
Db 904 CAGCTTGATGAGAAAACTCTGTACAGTGCAGCAGTGTAAAGATGGTTCCAGCTTCA 963
QY 976 AAGAGGTTCACTATCCATPAGATCTCTAATGTCTTACACTTTCTCTGAAACGTTTGGCA 1035
Db 964 AAGAGGTTCACTATCCATPAGATCTCTAATGTCTTACACTTTCTCTGAAACGTTTGGCA 1023
QY 1036 AATTTTACCGTGAAAAAATTTGCTAAGGATGTGAAATACCTGTAGTATCTTTGATATTCGG 1095
Db 1024 AATTTTACCGTGAAAAAATTTGCTAAGGATGTGAAATACCTGTAGTATCTTTGATATTCGG 1083
QY 1096 CCATATATGCTCAACCAAGCGAGGCCAATTTGTCTACAGTCTTGTATGCAAGTGTGCTC 1155
Db 1084 CCATATATGCTCAACCAAGCGAGGCCAATTTGTCTACAGTCTTGTATGCAAGTGTGCTC 1143
QY 1156 CACACTGGTTTTAAATTTGCCATGCTGGCCATTTCTGTCTACATAAAGCTAGCAATGGC 1215
Db 1144 CACACTGGTTTTAAATTTGCCATGCTGGCCATTTCTGTCTACATAAAGCTAGCAATGGC 1203
QY 1216 CTCTGGTATCAAAATGAATGATCCTCATTTGTATCTACAGTGTATTTAGATCGGTACTCAGC 1275
Db 1204 CTCTGGTATCAAAATGAATGATCCTCATTTGTATCTTACAGTGTATTTAGATCGGTACTCAGC 1263

QY 1276 CAAACAAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGAGGTGAACCTT 1335
Db 1264 CAAACAAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGAGGTGAACCTT 1323
QY 1336 ACTCATCCCAACCATCAGTCCCGGCCAGTCTCTCTCCCGGCCCGTCTCATCTAGTCAGCGGGTT 1395
Db 1324 ACTCATCCCAACCATCAGTCCCGGCCAGTCTCTCTCCCGGCCCGTCTCATCTAGTCAGCGGGTT 1383
QY 1396 GTCAACCAACAAACAGGCTGCGCCAGGCTTTTATCGGACCAACAGCTTCCTCTCACATGATA 1455
Db 1384 GTCAACCAACAAACAGGCTGCGCCAGGCTTTTATCGGACCAACAGCTTCCTCTCACATGATA 1443
QY 1456 AAGAAATCCACTCACTTAAATGGGACTGCAACATTGAAAGACACGCAAGCAGTTCCTCATG 1515
Db 1444 AAGAAATCCACTCACTTAAATGGGACTGCAACATTGAAAGACACGCAAGCAGTTCCTCATG 1503
QY 1516 TCGAGTCCCTAAACGGGAATTCAGAGTGTCAACAGGGCTAGTCTCTGTTAATGCTTCAGCTTCT 1575
Db 1504 TCGAGTCCCTAAACGGGAATTCAGAGTGTCAACAGGGCTAGTCTCTGTTAATGCTTCAGCTTCT 1563
QY 1576 GTCCAAAACCTGCTCAGTCTAATAGGTCTCTCAGTGTATCCAGAGACATCCTAAGAAACAAAA 1635
Db 1564 GTCCAAAACCTGCTCAGTCTAATAGGTCTCTCAGTGTATCCAGAGACATCCTAAGAAACAAAA 1623
QY 1636 ATTAACAATCAGTATTCACAACAAGTTCCTGCTGTGCGCAGTGTCACTCAACTAACCTT 1695
Db 1624 ATTAACAATCAGTATTCACAACAAGTTCCTGCTGTGCGCAGTGTCACTCAACTAACCTT 1683
QY 1696 CATAGTAAATCTTTTCGGAGAACCTTACCAAGCCGTTCCCTCTTCTACGATTAACCAATTC 1755
Db 1684 CATAGTAAATCTTTTCGGAGAACCTTACCAAGCCGTTCCCTCTTCTTACCATTACCAATTC 1743
QY 1756 CGAGTACAGTCTTACCTCGAAACGATCTACGATGTCACTTCTAGTAAAGTAAACAAACCG 1815
Db 1744 CGAGTACAGTCTTACCTCGAAACGATCTACGATGTCACTTCTAGTAAAGTAAACAAACCG 1803
QY 1816 ATCCCCCGCAGTGAATCTCTGCTCCAGGCCCGTGTGAATGGCAAAATCCAAAGCTGAACCTC 1875
Db 1804 ATCCCCCGCAGTGAATCTCTGCTCCAGGCCCGTGTGAATGGCAAAATCCAAAGCTGAACCTC 1863
QY 1876 AGCGTGTGTTGCCCTATGGCGCCGAGTCTCTCTGAGGACTCTGACGAGAGTCAAAAGGG 1935
Db 1864 AGCGTGTGTTGCCCTATGGCGCCGAGTCTCTCTGAGGACTCTGACGAGAGTCAAAAGGG 1923
QY 1936 CTGGCAAGAGAGAAATGGGATTTGTTACGATTTCTGAGTCTCCACTCTCCCGGCCAAGATGCC 1995
Db 1924 CTGGCAAGAGAGAAATGGGATTTGTTACGATTTGTTAGTCTCCCACTCTCCCGGCCAAGATGCC 1983
QY 1996 GAAGATGAGGAGGCACTCCGCAAGAGCTTCAAGAACCCCATGACCCCTAAAACGGTCTAAT 2055
Db 1984 GAAGATGAGGAGGCACTCCGCAAGAGCTTCAAGAACCCCATGACCCCTAAAACGGTCTAAT 2043
QY 2056 AGTCGACACGCGACGTGACCCGAAAGAAAACGGCTAGCGCTGTATGGTGCAGCTGC 2115
Db 2044 AGTCGACACGCGACGTGACCCGAAAGAAAACGGCTAGCGCTGTATGGTGCAGCTGC 2103
QY 2116 CAAGGCCAGCTGCTGCTCACTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTTGA 2175
Db 2104 CAAGGCCAGCTGCTGCTCACTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTTGA 2163
QY 2176 AAGTTGATGCTGCTCTCTTTGCTGTCTCTCCAGAGACAAAAATCTTAGAGACCTTCAGG 2235
Db 2164 AAGTTGATGCTGCTCTCTTTGCTGTCTCTCCAGAGACAAAAATCTTAGAGACCTTCAGG 2223
QY 2236 CTTAGCAACAACTGAAGGCTGACCGATGAATGATGTCACTGGAGCAGAGAGGGG 2295
Db 2224 CTTAGCAACAACTGAAGGCTGACCGATGAATGATGTCACTGGAGCAGAGAGGGG 2283
QY 2296 CCTCCGAGGACCGGACGCGGAGCTCAGCTGGAGCCCGCCCGCGAATCCCTGGAG 2355
Db 2284 CCTCCGAGGACCGGACGCGGAGCTCAGCTGGAGCCCGCCCGCGAATCCCTGGAG 2343

QY	2356	GAGC	2359	QY	61	CAGAGTTGAA	CAATGACCATAGTTGCA	CAAAAGCTTCTGAATCTTCA	AGCCATCAGCCTAT	120
DB	2344	GAGC	2347	DB	179	CAGAGTTGAA	CAATGACCATAGTTGCA	CAAAAGCTTCTGAATCTTCA	AGCCATCAGCCTAT	238
QY				QY	121	CAGAAATCAG	CGCTGGCAGCTCCGAGG	CAGTCTCACCTGGAGACAT	GGATCGAGGTTCTGCC	180
DB				DB	239	CAGAAATCAG	CGCTGGCAGCTCCGAGG	CAGTCTCACCTGGAGACAT	GGATCGAGGTTCTGCC	298
QY				QY	181	AGCTGGGTG	CTGTCTTCTTCAATGAT	GTGTCAAATCACA	CACTTTCTTTTAGGACCA	240
DB				DB	299	AGCTGGGTG	CTGTCTTCTTCAATGAT	GTGTCAAATCACA	CACTTTCTTTTAGGACCA	358
QY				QY	241	GTACCTGGT	GTCTGTATTTTATTCG	AGTTTCATCTGTA	CTGATAAATCAAAACCATC	300
DB				DB	359	GTACCTGGT	GTCTGTATTTTATTCG	AGTTTCATCTGTA	CTGATAAATCAAAACCATC	418
QY				QY	301	CAAAAGGAT	CAAGCCTTAGTGATG	CGTCCTCCACAGAAAGT	CTTTTCCCATCT	360
DB				DB	419	CAAAAGGAT	CAAGCCTTAGTGATG	CGTCCTCCACAGAAAGT	CTTTTCCCATCT	478
QY				QY	361	GAGAAATTT	GTCTTAAGTGGCAACAA	AACTCATAGAGTTGG	AGCTGGGCTCCAGAAATTTG	420
DB				DB	479	GAGAAATTT	GTCTTAAGTGGCAACAA	AACTCATAGAGTTGG	AGCTGGGCTCCAGAAATTTG	538
QY				QY	421	GGCAATAC	CTGTTTGGCAATG	CAGCCTGCAGTGT	TTTAACTACACACACCTCT	480
DB				DB	539	GGCAATAC	CTGTTTGGCAATG	CAGCCTGCAGTGT	TTTAACTACACACACCTCT	598
QY				QY	481	AATTACAT	CTGTATCATGAACT	CCAAACATGTCATG	CGAAGGCTTTTGTATGATG	540
DB				DB	599	AATTACAT	CTGTATCATGAACT	CCAAACATGTCATG	CGAAGGCTTTTGTATGATG	658
QY				QY	541	TGTACAAT	GCAAGCACATATTACC	CAGGCACCTCAGTA	ATCTCGGGACCGTTATTAAACCA	600
DB				DB	659	TGTACAAT	GCAAGCACATATTACC	CAGGCACCTCAGTA	ATCTCGGGACCGTTATTAAACCA	718
QY				QY	601	ATGTTTGT	CATCAATGAGATG	CGCGCTATAGTAGG	CACCTCGTTTGGAAACCAAGAA	660
DB				DB	719	ATGTTTGT	CATCAATGAGATG	CGCGCTATAGTAGG	CACCTCGTTTGGAAACCAAGAA	778
QY				QY	661	GATGCCCAT	GAATTCCTTCATAC	ACTGTTGATGCTAT	GCAGAAAGCATCTTTGAATGGC	720
DB				DB	779	GATGCCCAT	GAATTCCTTCATAC	ACTGTTGATGCTAT	GCAGAAAGCATCTTTGAATGGC	838
QY				QY	721	AGCAATAAA	TTAGACAGACACACCC	AGGCCACCACTCTT	CTTTGTTCAGATATTTGGAGGA	780
DB				DB	839	AGCAATAAA	TTAGACAGACACACCC	AGGCCACCACTCTT	CTTTGTTCAGATATTTGGAGGA	898
QY				QY	781	TACCTAAG	ATCTAGAGTCAAAAT	GTTTAAATTGCAAG	GGCGTTTTCAGATACCTTTGATCCA	840
DB				DB	899	TACCTAAG	ATCTAGAGTCAAAAT	GTTTAAATTGCAAG	GGCGTTTTCAGATACCTTTGATCCA	958
QY				QY	841	TATCTTGA	TATACATTTGGAGATA	AGGCTGCTCAGAGT	GTCAACAGGCAATTTGGAGCAG	900
DB				DB	959	TATCTTGA	TATACATTTGGAGATA	AGGCTGCTCAGAGT	GTCAACAGGCAATTTGGAGCAG	1018
QY				QY	901	TTTGTGA	AGCGGACACAGCTTT	GTATGGAGAAAAC	TCTGTAAGTGCAGCAAGTGTAA	960
DB				DB	1019	TTTGTGA	AGCGGACACAGCTTT	GTATGGAGAAAAC	TCTGTAAGTGCAGCAAGTGTAA	1078
QY				QY	961	ATGTTTCC	AGCTTCAAGAGGTT	CACATATCAGAT	ATCTTCTTACACTTTCT	1020
DB				DB	1079	ATGTTTCC	AGCTTCAAGAGGTT	CACATATCAGAT	ATCTTCTTACACTTTCT	1138
QY				QY	1021	CTGAAAG	GTGTTGCAAAATTTT	TACCGTGGAAAAT	TCTTAAGGATGTGAATACCTCTGAG	1080
DB				DB	1139	CTGAAAG	GTGTTGCAAAATTTT	TACCGTGGAAAAT	TCTTAAGGATGTGAATACCTCTGAG	1198
QY				QY	1081	TATCTTGA	TATTCGCAATATAT	GTCTCAACCCAC	CGAGAGCCAAATGTCTACGCTTTG	1140
DB				DB	1199	TATCTTGA	TATTCGCAATATAT	GTCTCAACCCAC	CGAGAGCCAAATGTCTACGCTTTG	1258
QY				QY	1141	TATCAGT	GTGGTCCACACTGT	TTTAAATGGCC	ATTACTTCTGCTACATA	1200

RESULT 6

ABZ11508

ID ABZ11508 standard; cDNA; 4483 BP.

AC

ABZ11508;

DT 20-JAN-2003 (first entry)

DE Human polynucleotide SEQ ID NO 390.

Human; genome mapping; gene therapy; food supplement; virus; fungus; cell-proliferative disorder; neurodegenerative disease; bacterial; Parkinson's disease; Alzheimer's disease; autoimmune disease; multiple sclerosis; diabetes; genetic disorder; wound; burn; infection; arthritis; cytostatic; immunomodulator; nootropic; neuroprotective; antiparkinsonian; antidiabetic; immunosuppressive; dermatological; haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide; antiarthritic; gene; ss.

OS Homo sapiens.

XN WO200270539-A2.

XX 12-SEP-2002.

XX 05-MAR-2002; 2002WO-US005095.

XX 05-MAR-2001; 2001US-00799451.

XX (HYSEQ-) HYSEQ INC.

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2002-759812/82.
DR P-PSDB; ABP69291.

XX New polynucleotides comprising sequences assembled from expressed
PT sequence tags (ESTs), useful for treating cell-proliferative,
PT neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT or coagulation disorders.

XX Claim 1; SEQ ID NO 390; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
CC nucleotide sequence selected from any of 948 sequences (ABZ119-
CC ABZ12066) or their mature protein coding portion, active domain coding
CC protein or complementary sequences. The polynucleotides are useful for
CC identifying expressed genes or for physical mapping of human genome. The
CC encoded polypeptides (ABP68902-ABP6849) are useful as molecular weight
CC markers, as a food supplement, for generating antibodies, in medical
CC imaging, screening and diagnostic assays and for treating cell-
CC proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC arthritis, etc. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WPI at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 4483 BP; 1127 A; 1270 C; 1169 G; 917 T; 0 U; 0 Other;

Query Match 95.8%; Score 2310; DB 6; Length 4483;

Best Local Similarity 99.8%; Pred. NO. 0;

Matches 2313; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

CC diseases, dystonia, multiple sclerosis, rheumatoid arthritis, psoriasis,
CC cataracts, restenosis, atherosclerosis, inflammation, skin disorders,
CC glaucoma, obesity, muscular dystrophy, AIDS, infertility, cardiovascular
CC disease, coagulation disease, ischemia, hypertension, asthma, immune
CC disease, epilepsy, angina, neurodegeneration, diabetes, anxiety,
CC depression, schizophrenia, viral disease, gastric ulcers, stroke,
CC Alzheimer's disease and as a contraceptive
XX
SQ Sequence 1679 BP; 461 A; 443 C; 378 G; 393 T; 0 U; 4 Other;

Query Match 69.5%; Score 1675.6; DB 6; Length 1679;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1675; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 694 GCTATGAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACCCAGGCCACC 753
Db 1 GCTATGAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACCCAGGCCACC 60

QY 754 ACTCTTGTGTGTCAGATATTGAGAGATACCTAAGATCTAGAGTCAATGTTTAAATTGC 813
Db 61 ACTCTTGTGTGTCAGATATTGAGAGATACCTAAGATCTAGAGTCAATGTTTAAATTGC 120

QY 814 AAGGCCGTTTCAGATATTGATCCATATCTTCATATAACATTGGAGATAAAGGCTGCT 873
Db 121 AAGGCCGTTTCAGATATTGATCCATATCTTCATATAACATTGGAGATAAAGGCTGCT 180

QY 874 CAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCGGAAACAGCTTGATGGAGAAAC 933
Db 181 CAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCGGAAACAGCTTGATGGAGAAAC 240

QY 934 TCGTACAAAGTCAGCAAGTGTAAAGATGTTCCAGCTTCAAGAGGTTCACTATCCAT 993
Db 241 TCGTACAAAGTCAGCAAGTGTAAAGATGTTCCAGCTTCAAGAGGTTCACTATCCAT 300

QY 994 AGATCCCTAATGTTCTTACACTTCTCTGAAACGTTTTGCAAAATTTTACCGGTGGAATA 1053
Db 301 AGATCCCTAATGTTCTTACACTTCTCTGAAACGTTTTGCAAAATTTTACCGGTGGAATA 360

QY 1054 ATTGTAAGGATGTGAATACCTGAGTATCTTGATATTCGGCCATATATGTCTCAACC 1113
Db 361 ATTGTAAGGATGTGAATACCTGAGTATCTTGATATTCGGCCATATATGTCTCAACC 420

QY 1114 AACGAGAGCCAAATGTCTAAGTGTGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1173
Db 421 AACGAGAGCCAAATGTCTAAGTGTGATGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480

QY 1174 CATGCTGCCCATTACTTCTGCTACATAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCT 1233
Db 481 CATGCTGCCCATTACTTCTGCTACATAAAGCTAGCAATGGCTCTGCTGCTGCTGCTGCTGCTGCT 540

QY 1234 GACTCCATTGATCTACCAAGTATATTAGATCGGTACTCAGCCAAAGCCCTATGTGCTC 1293
Db 541 GACTCCATTGATCTACCAAGTATATTAGATCGGTACTCAGCCAAAGCCCTATGTGCTC 600

QY 1294 TTTTATATCAGTCCCATGATGTGAAATATGAGGTGAATCTACTCATCCACCCATAGC 1353
Db 601 TTTTATATCAGTCCCATGATGTGAAATATGAGGTGAATCTACTCATCCACCCATAGC 660

QY 1354 CCGGCCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGTCAACCAACAAACAGGCT 1413
Db 661 CCGGCCAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGTCAACCAACAAACAGGCT 720

QY 1414 GCGCCAGGCTTTATCGGACACAGCTTCCCTCTCATGATTAAGATCCACCTCACTTA 1473
Db 721 GCGCCAGGCTTTATCGGACACAGCTTCCCTCTCATGATTAAGATCCACCTCACTTA 780

QY 1474 AATGGACTGGACCATTTGAAGACAGCCAGGATTCATGTCGAGTCCCTAACCGGAAT 1533
Db 781 AATGGACTGGACCATTTGAAGACAGCCAGGATTCATGTCGAGTCCCTAACCGGAAT 840

QY 1534 TCCAGTGTCAACAGGCTAGTCTCTGTTTAAATGCTTCTGATCCAAATCTGGTCAGTT 1593
Db 841 TCCAGTGTCAACAGGCTAGTCTCTGTTTAAATGCTTCTGATCCAAATCTGGTCAGTT 900

QY 1594 AATAGTCTCCTCAGTGTATCCAGAAACATCTTAAGAAAACAAAATTAATCAATCAGTATTAC 1653
Db 901 AATAGTCTCCTCAGTGTATCCAGAAACATCTTAAGAAAACAAAATTAATCAATCAGTATTAC 960

QY 1654 AACAAAGTTCGCTGTTCCGAGTGTGATCTCAACCTTAACCTTCAATAGTAAATTTTGGAG 1713
Db 961 AACAAAGTTCGCTGTTCCGAGTGTGATCTCAACCTTCAATAGTAAATTTTGGAG 1020

QY 1714 AACCTTACCAAGCCGTTCCCTCTCTTACCATTACCAATCTTCAGTACAGTCTACCTCG 1773
Db 1021 AACCTTACCAAGCCGTTCCCTCTCTTACCATTACCAATCTTCAGTACAGTCTACCTCG 1080

QY 1774 AAGCATCTACGATGTGTCAGTCTTAGTAAAGTAAACAAACCGATCCCGCAGTGAATCC 1833
Db 1081 AAGCATCTACGATGTGTCAGTCTTAGTAAAGTAAACAAACCGATCCCGCAGTGAATCC 1140

QY 1834 TGCTCCAGCCGTTGATGATGGAATCCAGCTGAATCCAGCTGAATCCAGCTGCTGCTGCTAT 1893
Db 1141 TGCTCCAGCCGTTGATGATGGAATCCAGCTGAATCCAGCTGCTGCTGCTGCTAT 1200

QY 1894 GCGCCAGGCTCTCTGAGGACTCTGACGAGAGTCAAGGGGCTGGCAAGGAGAAATGG 1953
Db 1201 GCGCCAGGCTCTCTGAGGACTCTGACGAGAGTCAAGGGGCTGGCAAGGAGAAATGG 1260

QY 1954 ATTGTTACGATTTGAGTCTCCACTCTCCGCGCAAGATGCCGAAGATGAGGAGGCCACT 2013
Db 1261 ATTGTTACGATTTGAGTCTCCACTCTCCGCGCAAGATGCCGAAGATGAGGAGGCCACT 1320

QY 2014 CCGCAGAGCTTCAAGAACCCATGACCTTAACCGTCTTAATAGTGCAGACAGCGACAGT 2073
Db 1321 CCGCAGAGCTTCAAGAACCCATGACCTTAACCGTCTTAATAGTGCAGACAGCGACAGT 1380

QY 2074 GACCCGAGAGAAAGCGCTAGCGCTGATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2133
Db 1381 GACCCGAGAGAAAGCGCTAGCGCTGATGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1440

QY 2134 CACTCAGAAAAATCCCTTTGCTTAAAGCAAAACCGTCTTCTGGAAGTGTGATGCTGCTCT 2193
Db 1441 CACTCAGAAAAATCCCTTTGCTTAAAGCAAAACCGTCTTCTGGAAGTGTGATGCTGCTCT 1500

QY 2194 TTGCTGTCTCTCCAGAAAGACAAATCTTAGAGCTTTAGAGCTTTAGCAAACTGAAA 2253
Db 1501 TTGCTGTCTCTCCAGAAAGACAAATCTTAGAGCTTTAGAGCTTTAGCAAACTGAAA 1560

QY 2254 GGCTCGAGGATGAATCAGTGCACCTGAGCAGAGAGGGGCTTCCGAGGACCGCGAC 2313
Db 1561 GGCTCGAGGATGAATCAGTGCACCTGAGCAGAGAGGGGCTTCCGAGGACCGCGAC 1620

QY 2314 GCGGAGCCTCAGCTTGGCAGCCCGCCCGCGAATCCCTGAGGAGCAGATCGGCGCG 2372
Db 1621 GCGGAGCCTCAGCTTGGCAGCCCGCCCGCGAATCCCTGAGGAGCAGATCGGCGCG 1679

RESULT 8

AAK79798

ID AAK79798 standard; DNA; 16682 BP.

XX AAK79798;

AC AAK79798;

XX 07-NOV-2001 (first entry)

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34610.

DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

XX cytostatic; gene therapy; vaccine; metastasis; ds.

KW Homo sapiens.

XX OS WO200157182-A2.

XX PN 09-AUG-2001.

XX PD

XX XX

PF 17-JAN-2001; 2001WO-US001354.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-01981123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215113P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226968P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 21-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.

PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 13-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0244617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,

PT useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 34610; 3071pp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AM82170 to AM91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AM82169
CC represent sequences used in the exemplification of the present invention
XX
XX
SQ Sequence 16682 BP; 4217 A; 3754 C; 4167 G; 4538 T; 0 U; 6 Other;
Query Match 30.1%; Score 727.2; DB 4; Length 16682;
Best Local Similarity 98.3%; Pred. No. 9e-204;
Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1435 CAGCTCCCTCTCACATGATGAAGATCCACTCACTTAATGGACCTGGACCATTTGAAA 1494
DB 9429 CAGGTGCGCTCTGCTTACTTCAGAAATCACTCACTTAATGGACCTGGACCATTTGAAA 9488
QY 1495 GACAGCCCAAGCAGTTCATGTCGAGTCCTAAACGGGAATTCAGGTGCAACAGGGCTAGT 1554
DB 9489 GACAGCCCAAGCAGTTCATGTCGAGTCCTAAACGGGAATTCAGGTGCAACAGGGCTAGT 9548
QY 1555 CCGTGTAAATGCTTACGTTCTGTCGAAACTGCTCAGTATATAGCTCTCAGTATCCCA 1614
DB 9549 CCGTGTAAATGCTTACGTTCTGTCGAAACTGCTCAGTATATAGCTCTCAGTATCCCA 9608
QY 1615 GAACATCTTGAAGAAACAAAATTAACATCAGTATTCACAAAGTTCGCTTTCGCCAG 1674
DB 9609 GAACATCTTGAAGAAACAAAATTAACATCAGTATTCACAAAGTTCGCTTTCGCCAG 9668
QY 1675 TGTCAGTCTCAACCTTCACTAGTAAATCTTTGGAGAACCTTCAAGCCGCTTCC 1734
DB 9669 TGTCAGTCTCAACCTTCACTAGTAAATCTTTGGAGAACCTTCAAGCCGCTTCC 9728
QY 1735 TCTTCTACATTAACCAATTCGTCAGTACAGTCTCTCGAAGCAGCATCTACGATGTCAGTT 1794
DB 9729 TCTTCTACATTAACCAATTCGTCAGTACAGTCTCTCGAAGCAGCATCTACGATGTCAGTT 9788
QY 1795 TCTAGTAAAGTAAACAAACCGATCCCGCAGTGAATCTCGTCCAGCCCGTGATGAAT 1854
DB 9789 TCTAGTAAAGTAAACAAACCGATCCCGCAGTGAATCTCGTCCAGCCCGTGATGAAT 9848
QY 1855 GGCATAATCAAGCTGAACCTCCAGCGTGTGGTCCCTATGGCCCGCAGTCTCTGAGGAC 1914
DB 9849 GGCATAATCAAGCTGAACCTCCAGCGTGTGGTCCCTATGGCCCGCAGTCTCTGAGGAC 9908
QY 1915 TCTGACGAGGATCAAGGGGCTGGCAGAGGAATGGATGCTGATGTCAGTGTGAGCTCC 1974
DB 9909 TCTGACGAGGATCAAGGGGCTGGCAGAGGAATGGATGCTGATGTCAGTGTGAGCTCC 9968
QY 1975 CACTCTCCCGCCCAAGATGCCAAGATGAGGAGGCCCACTCCCGCAGCGTTCGAAGAACCC 2034
DB 9969 CACTCTCCCGCCCAAGATGCCAAGATGAGGAGGCCCACTCCCGCAGCGTTCGAAGAACCC 10028
QY 2035 ATGACCTTAAACGGTGTCTAATAGTGACAGCAGCAGTACCCGGAAGAAACGGCCTA 2094
DB 10029 ATGACCTTAAACGGTGTCTAATAGTGACAGCAGCAGTACCCGGAAGAAACGGCCTA 10088
QY 2095 GCGCCTGATGTCGCCAGCTGCCAAGGCCAGCTCCCTGCACTCAGAAAAATCCCTTTGCT 2154

DB 10089 GCGCCTGATGTCGCCAGCTGCCAAGGCCAGCTCCCTGCACTCAGAAAAATCCCTTTGCT 10148
QY 2155 AAGGCAACGGTCTTCTCGAAGAGTTGA 2182
DB 10149 AAGGCAACGGTCTTCTCGAAGAGTTGA 10176
RESULT 9
ID AAK70010
AC AAK70010; DNA: 16682 BP.
XX
XX 06-NOV-2001 (first entry)
DT Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24822.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
PN 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.

```

PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231245P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234055P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 26-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239355P.
PR 20-OCT-2000; 2000US-0239377P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 08-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 08-SEP-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 06-DEC-2000; 2000US-0256719P.
PR 08-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0254097P.
PR 05-JAN-2001; 2001US-0254967P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
XX useful for preventing, diagnosing and/or treating cancers and metastasis.
XX Disclosure; SEQ ID NO 24822; 307lpp + Sequence Listing; English.
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
XX amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
XX activity, and can be used in gene therapy and vaccine production. (I)
XX proteins and polynucleotides may be used in the prevention, diagnosis and
XX treatment of diseases associated with inappropriate (I) expression. For
XX example, they may be used to treat disorders associated with decreased
XX expression by rectifying mutations or deletions in a patient's genome
XX that affect the activity of (I) by expressing inactive proteins or to
XX supplement the patients own production of (I). Additionally, (I)
XX polynucleotides may be used to produce the secreted (I), by inserting the
XX nucleic acids into a host cell and culturing the cell to express the
XX protein. (I) proteins and polynucleotides may be used to prevent,
XX diagnose and treat immune/haematopoietic-related diseases, especially
XX cancers and cancer metastases of haematopoietic-derived cells. AAK64703
XX to AAK87694 represent human immune/haematopoietic antigen genomic
XX sequences from the present invention. AAK54942 to AAK54950 and AAK82169
XX represent sequences used in the exemplification of the present invention
XX SQ Sequence 16682 BP; 4217 A; 3754 C; 4167 G; 4538 T; 0 U; 6 Other;
Query Match 30.1%; Score 727.2; DB 4; Length 16682;
Best Local Similarity 98.3%; Pred No. 9e-204;
Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
QY 1435 CAGCTTCCCTCTCAGATGATTAAGATCCACTCCTTAATGGGACTGGACCTTGAAT 1494
Db 9429 CAGGTGCGCTCTGCTTACTTCCAGATCCACTCCTTAATGGGACTGGACCTTGAAT 9488
QY 1495 GACACGCCAAGCAGTTCATGCTCAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGT 1554
Db 9489 GACACGCCAAGCAGTTCATGCTCAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGT 9548
QY 1555 CCTGTTAATGCTTCAGCTTCTGTCACAACTGGTCAGTTAATAGTCTCAGTGTATCCCA 1614
Db 9549 CCTGTTAATGCTTCAGCTTCTGTCACAACTGGTCAGTTAATAGTCTCAGTGTATCCCA 9608
QY 1615 GAACATCCTAAGAAACAAAAAATTACATCAGTATTACACAAAGTTGCTGTTGCGCAG 1674
Db 9609 GAACATCCTAAGAAACAAAAAATTACATCAGTATTACACAAAGTTGCTGTTGCGCAG 9668
QY 1675 TGTCAGTCTCAACTTAACCTTCATAGTAATCTTTGAGAACCCCTACCAAGCCCGTTCCC 1734

```

Wed Aug 18 13:53:01 2004

us-10-049-745-31.rng

```
Db      9669  TGTGAGTCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTTACCAGCCCGTTCCC 9728
Qy      1735  TCTTCTACCAATTACCAATTTCTGAGTACAGTCTTACCTCGAAGCAATCTACGATGTGAGTT 1794
Db      9729  TCTTCTACCAATTACCAATTTCTGAGTACAGTCTTACCTCGAAGCAATCTACGATGTGAGTT 9788
Qy      1795  TCTAGTAAAGTAACAACCGATCCCGCGAGTGAATCTCTGCTCCGAGCCCGTGATGAAT 1954
Db      9789  TCTAGTAAAGTAACAACCGATCCCGCGAGTGAATCTCTGCTCCGAGCCCGTGATGAAT 9848
Qy      1855  GGCATATCCAAAGCTGAACCTCCAGCTGCTGGTGCCTCTATGGCGCCGAGTCTCTCTGAGGAC 1914
Db      9849  GGCATATCCAAAGCTGAACCTCCAGCTGCTGGTGCCTCTATGGCGCCGAGTCTCTCTGAGGAC 9908
Qy      1915  TCTGACGAGGAGTCAAGGGGCTGGGCAAGAGATGGGATGGTACGATGTGAGTCTCC 1974
Db      9909  TCTGACGAGGAGTCAAGGGGCTGGGCAAGAGATGGGATGGTACGATGTGAGTCTCC 9968
Qy      1975  CACTCTCCCGGCAAGATGCGGAAGATGAGAGGCCCACTCGGACGAGCTTCAAGAACCC 2034
Db      9969  CACTCTCCCGGCAAGATGCGGAAGATGAGAGGCCCACTCGGACGAGCTTCAAGAACCC 10028
Qy      2035  ATGACCTTAACGGTGTCTAATAGTGACAGAGCCAGTCAACCGAAGAAACCGGCTA 2094
Db      10029  ATGACCTTAACGGTGTCTAATAGTGACAGAGCCAGTCAACCGAAGAAACCGGCTA 10088
Qy      2095  GCGCTGATGGTGCAGTGTCCAGGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2154
Db      10089  GCGCTGATGGTGCAGTGTCCAGGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10148
Qy      2155  AAGCAAAACGGTCTTCTGGAAGTTGA 2182
Db      10149  AAGCAAAACGGTCTTCTGGAAGTTGA 10176

RESULT 10
AAK79799
ID   AAK79799 standard; DNA; 21436 BP.
XX
AC   AAK79799;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34611.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
XX   04-FEB-2000; 2000US-0180628P.
XX   24-FEB-2000; 2000US-0184664P.
XX   02-MAR-2000; 2000US-0186350P.
XX   16-MAR-2000; 2000US-0189874P.
XX   17-MAR-2000; 2000US-0190076P.
XX   18-APR-2000; 2000US-0198123P.
XX   19-MAY-2000; 2000US-0205515P.
XX   07-JUN-2000; 2000US-0209467P.
XX   28-JUN-2000; 2000US-0214886P.
XX   30-JUN-2000; 2000US-0215135P.
XX   07-JUL-2000; 2000US-0216647P.
XX   07-JUL-2000; 2000US-0216980P.
XX   11-JUL-2000; 2000US-0217487P.
XX   11-JUL-2000; 2000US-0217496P.
XX   14-JUL-2000; 2000US-0218290P.
XX   26-JUL-2000; 2000US-0220963P.
XX   26-JUL-2000; 2000US-0220963P.
XX   14-AUG-2000; 2000US-0224518P.
XX   14-AUG-2000; 2000US-0224519P.
XX   14-AUG-2000; 2000US-0225213P.
XX   14-AUG-2000; 2000US-0225214P.
XX   14-AUG-2000; 2000US-0225266P.
XX   14-AUG-2000; 2000US-0225267P.
XX   14-AUG-2000; 2000US-0225270P.
XX   14-AUG-2000; 2000US-0225447P.
XX   14-AUG-2000; 2000US-0225757P.
XX   14-AUG-2000; 2000US-0225758P.
XX   14-AUG-2000; 2000US-0225759P.
XX   18-AUG-2000; 2000US-0226279P.
XX   22-AUG-2000; 2000US-0226681P.
XX   22-AUG-2000; 2000US-0226868P.
XX   22-AUG-2000; 2000US-0227182P.
XX   22-AUG-2000; 2000US-0227009P.
XX   30-AUG-2000; 2000US-0228924P.
XX   01-SEP-2000; 2000US-0229287P.
XX   01-SEP-2000; 2000US-0229343P.
XX   01-SEP-2000; 2000US-0229344P.
XX   01-SEP-2000; 2000US-0229345P.
XX   05-SEP-2000; 2000US-0229509P.
XX   05-SEP-2000; 2000US-0229513P.
XX   06-SEP-2000; 2000US-0230437P.
XX   06-SEP-2000; 2000US-0230438P.
XX   08-SEP-2000; 2000US-0231242P.
XX   08-SEP-2000; 2000US-0231243P.
XX   08-SEP-2000; 2000US-0231244P.
XX   08-SEP-2000; 2000US-0231413P.
XX   08-SEP-2000; 2000US-0231414P.
XX   08-SEP-2000; 2000US-0232080P.
XX   08-SEP-2000; 2000US-0232081P.
XX   12-SEP-2000; 2000US-0231968P.
XX   14-SEP-2000; 2000US-0232397P.
XX   14-SEP-2000; 2000US-0232398P.
XX   14-SEP-2000; 2000US-0232399P.
XX   14-SEP-2000; 2000US-0232400P.
XX   14-SEP-2000; 2000US-0232401P.
XX   14-SEP-2000; 2000US-0233063P.
XX   14-SEP-2000; 2000US-0233064P.
XX   14-SEP-2000; 2000US-0233065P.
XX   21-SEP-2000; 2000US-0234223P.
XX   21-SEP-2000; 2000US-0234274P.
XX   25-SEP-2000; 2000US-0234997P.
XX   25-SEP-2000; 2000US-0234998P.
XX   26-SEP-2000; 2000US-0235484P.
XX   27-SEP-2000; 2000US-0235834P.
XX   27-SEP-2000; 2000US-0236327P.
XX   29-SEP-2000; 2000US-0236367P.
XX   29-SEP-2000; 2000US-0236368P.
XX   29-SEP-2000; 2000US-0236369P.
XX   29-SEP-2000; 2000US-0236370P.
XX   02-OCT-2000; 2000US-0236802P.
XX   02-OCT-2000; 2000US-0237037P.
XX   02-OCT-2000; 2000US-0237038P.
XX   02-OCT-2000; 2000US-0237039P.
XX   13-OCT-2000; 2000US-0239933P.
XX   13-OCT-2000; 2000US-0239937P.
XX   20-OCT-2000; 2000US-0240960P.
XX   20-OCT-2000; 2000US-0241221P.
XX   20-OCT-2000; 2000US-0241785P.
XX   20-OCT-2000; 2000US-0241786P.
XX   20-OCT-2000; 2000US-0241787P.
XX   20-OCT-2000; 2000US-0241808P.
XX   20-OCT-2000; 2000US-0241809P.
XX   01-NOV-2000; 2000US-0241826P.
XX   08-NOV-2000; 2000US-0244617P.
XX   08-NOV-2000; 2000US-0246474P.
XX   08-NOV-2000; 2000US-0246475P.
```

PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 08-NOV-2000; 2000US-0246707P.
 PR 08-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI, 2001-483426/52.
 Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 useful for preventing, diagnosing and/or treating cancers and metastasis.
 Disclosure; SEQ ID NO 34611; 307pp + Sequence Listing; English.
 AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 activity, and can be used in gene therapy and vaccine production. (I)
 proteins and polynucleotides may be used in the prevention, diagnosis and
 treatment of diseases associated with inappropriate (I) expression. For
 example, they may be used to treat disorders associated with decreased
 expression by rectifying mutations or deletions in a patient's genome
 that affect the activity of (I) by expressing inactive proteins or to
 supplement the patient's own production of (I). Additionally, (I)
 polynucleotides may be used to produce the secreted (I), by inserting the
 nucleic acids into a host cell and culturing the cell to express the
 protein. (I) proteins and polynucleotides may be used to prevent,
 diagnose and treat immune/haematopoietic-related diseases, especially
 cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 to AAK87694 represent human immune/haematopoietic antigen genomic

CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention
 XX
 SQ Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;
 Query Match 30.1%; Score 727.2; DB 4; Length 21436;
 Best Local Similarity 98.3%; Pred. No. 1e-203;
 Matches 735; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
 QY 1435 CAGCTTCCCTCTCACATGATAAAGATCCACCTCACTTAATCGGACTGGACCATTCGAAA 1494
 |||||
 Db 9432 CAGGTGGCTCTGCTTACTTCCAGAAATCCACCTCACTTAATCGGACTGGACCATTCGAAA 9491
 |||||
 QY 1495 GACACGCCAAGAGTTCCATGTCAGTCTCTAACGGGAAATTCACAGTGTCAACAGGGTAGT 1554
 |||||
 Db 9492 GACACGCCAAGAGTTCCATGTCAGTCTCTAACGGGAAATTCACAGTGTCAACAGGGTAGT 9551
 |||||
 QY 1555 CTTGTTAATGCTTCAGCTTCTGTCGAAAATGCTGAGTTAATAGTCTCTCAGTGATCCCA 1614
 |||||
 Db 9552 CTTGTTAATGCTTTCAGCTTCTGTCGAAAATGCTGAGTTAATAGTCTCTCAGTGATCCCA 9611
 |||||
 QY 1615 GAACATCTTAAGAAACAAAAAATTACAATCAGTATTCAACAAGAGTTGCTGTTCGCCAG 1674
 |||||
 Db 9612 GAACATCTTAAGAAACAAAAAATTACAATCAGTATTCAACAAGAGTTGCTGTTCGCCAG 9671
 |||||
 QY 1675 TGTGAGTCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCC 1734
 |||||
 Db 9672 TGTGAGTCTCAACCTTAACCTTCATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCC 9731
 |||||
 QY 1735 TCTTCTACCAATTACCAATTTCTGAGTACAGTCTACTCTGAAACGATCTACGATGTGAGTT 1794
 |||||
 Db 9732 TCTTCTACCAATTACCAATTTCTGAGTACAGTCTACTCTGAAACGATCTACGATGTGAGTT 9791
 |||||
 QY 1795 TCTAGTAAAGTAACAAACCCGATCCCGCAGTGAATCTGCTCCAGCCCGTGTGATGAAT 1854
 |||||
 Db 9792 TCTAGTAAAGTAACAAACCCGATCCCGCAGTGAATCTGCTCCAGCCCGTGTGATGAAT 9851
 |||||
 QY 1855 GGCAATCCAGCTGAATCCAGCGTGTGTGCCCTATGGCGCGAGTCTCTCTGAGGAC 1914
 |||||
 Db 9852 GGCAATCCAGCTGAATCCAGCGTGTGTGCCCTATGGCGCGAGTCTCTCTGAGGAC 9911
 |||||
 QY 1915 TCTGACGAGGAGTCAAGGGGCTGGGAGAGAGATGGATTTGGTACGATTTGTGAGTCC 1974
 |||||
 Db 9912 TCTGACGAGGAGTCAAGGGGCTGGGAGAGAGATGGATTTGGTACGATTTGTGAGTCC 9971
 |||||
 QY 1975 CACTCTCCCGCCAAAGATGCCGAAGTCAAGAGGCGACTCCGCAAGAGTTCACGAACCC 2034
 |||||
 Db 9972 CACTCTCCCGCCAAAGATGCCGAAGTCAAGAGGCGACTCCGCAAGAGTTCACGAACCC 10031
 |||||
 QY 2035 ATGACCTTAAACGGTGTAAATAGTGCACACAGCGACAGTGACCCGAAAGAAACGCGCTA 2094
 |||||
 Db 10032 ATGACCTTAAACGGTGTAAATAGTGCACACAGCGACAGTGACCCGAAAGAAACGCGCTA 10091
 |||||
 QY 2095 GCGCTGATGTGCGCAGCTGCCAGCCAGCTGCCCTGCCTGCACCTCAGAAAAATCCCTTGGT 2154
 |||||
 Db 10092 GCGCTGATGTGCGCAGCTGCCAGCCAGCTGCCCTGCCTGCACCTCAGAAAAATCCCTTGGT 10151
 |||||
 QY 2155 AAGCAAAACGGTCTTCTCTGGAAGTTGA 2182
 |||||
 Db 10152 AAGCAAAACGGTCTTCTCTGGAAGTTGA 10179
 |||||
 RESULT 11
 AAK70011
 ID AAK70011 standard; DNA; 21436 BP.
 XX
 AC AAK70011;
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24823.
 XX
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;


```
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0228287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249267P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0239678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465566/50.
XX P-PSDB; AAU23207.
XX Novel polypeptides and polynucleotides useful for diagnosing, preventing,
XX treating neural, immune system, muscular, reproductive, pulmonary,
XX cardiovascular, renal, proliferative disorders and cancerous diseases.
XX Claim 4; SEQ ID NO 303; 1180pp; English.
XX The present invention relates to the isolation of novel human enzyme
XX polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
XX encoding them. The enzyme polypeptides of the invention may comprise the
XX functional classes of oxidoreductases, transferases, hydrolases, lyases,
XX isomerases or ligases. The sequences of the invention are useful in the
XX diagnosis, treatment, prevention and/or prognosis of a wide range of
XX disorders including hyperproliferative disorders (e.g. cancer),
XX immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g.
XX arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic
XX disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma),
XX cardiovascular disorders (e.g. atherosclerosis), blood-related disorders
XX (e.g. haemophilia), reproductive disorders (e.g. infertility) and
XX infectious disorders (e.g. influenza). The polynucleotides of the
XX invention can also be used in gene therapy. AAS40785-AAS41684 represent
XX cDNA sequences encoding for the novel human enzyme polypeptides of the
XX invention. Note: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 1122 BP; 330 A; 247 C; 260 G; 285 T; 0 U; 0 Other;
SQ
Query Match 21.5%; Score 517.4; DB 4; Length 1122;
Best Local Similarity 97.1%; Pred. No. 3.1e-142; Indels 0; Gaps 0;
Matches 527; Conservative 0; Mismatches 16;
QY 1 GGGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 60
DB 12 GGGCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 71
```


PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250319P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-)	HUMAN GENOME SCI INC.
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-465566/50.	
DR	P-PSDB; AAU23743.	
XX		
PT	Novel polypeptides and polynucleotides useful for diagnosing, preventing,	
PT	treating neural, immune system, muscular, reproductive, pulmonary,	
PT	cardiovascular, renal, proliferative disorders and cancerous diseases.	
XX		
PS	Claim 4; SEQ ID NO 839; 1180pp; English.	
XX		
CC	The present invention relates to the isolation of novel human enzyme	
CC	polypeptides (AAU2915-AAU23814), and the cDNA and genomic sequences	
CC	encoding them. The enzyme polypeptides of the invention may comprise the	
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,	
CC	isomerases or ligases. The sequences of the invention are useful in the	
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of	

FN US6331427-B1.
 XX
 PD 18-DEC-2001.
 XX
 PF 26-MAR-1999; 99US-00280116.
 XX
 PR 26-MAR-1999; 99US-00280116.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Robison KE;
 XX
 DR WPI; 2002-129545/17.
 XX
 PT New polynucleotides encoding protease homologs of the G-protein-coupled
 PT protease family, useful in identifying agonists and antagonists for
 PT diagnosis and treatment of protease mediated disorders.
 XX
 PS Disclosure; Col 251-252; 245pp; English.
 XX
 CC The invention relates to an isolated human protease nucleic acid molecule
 CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
 CC defined in the specification. Also disclosed are production of an
 CC isolated polypeptide encoded by the nucleic acid, comprising introducing
 CC the nucleic acid into a host cell and culturing under conditions to
 CC express the protein from the nucleic acid, use of an antibody to detect
 CC the encoded protein in a sample and to modulate its in vivo activity,
 CC identifying agents that bind to the protein and identification of a
 CC polynucleotide agent that modulates the expression of the nucleic acid or
 CC its complement (i.e. gene therapy). The nucleic acid can be used to
 CC identify an agent that modulates the expression of the nucleic acid or
 CC nucleic acid, and can be used to isolate the protein. The nucleic acid
 CC can be used in diagnostic assays for determining nucleic acid expression
 CC as well as activity in the context of a biological sample (e.g., blood,
 CC serum, cells, tissue) to determine whether an individual has a disease or
 CC disorder, or is at risk of developing a disease or disorder, associated
 CC with aberrant expression or activity of the nucleic acid. The nucleic
 CC acid can be used to detect mutations in protease genes and gene
 CC expression products such as mRNA. The nucleic acid can be used as
 CC hybridisation probes to detect naturally-occurring genetic mutations in a
 CC protease gene. The nucleic acid can be used in drug screening methods to
 CC identify agonists and antagonists that can be used to diagnose and treat
 CC such protease mediated disorders e.g., proliferative, differentiative,
 CC developmental or haematopoietic disorders. The nucleic acid can be used
 CC as probes, primers, in biological assays, to determine patterns of gene
 CC expression, to design ribozymes and to construct transgenic animals. The
 CC present sequence represents one of the 268 disclosed human G-protein-
 CC coupled protease cDNA sequences
 XX
 SQ Sequence 722 BP; 189 A; 186 C; 145 G; 196 T; 0 U; 6 Other;
 Query Match 19.7%; Score 474.6; DB 6; Length 722;
 Best Local Similarity 92.9%; Pred. No. 1.2e-129;
 Matches 562; Conservative 0; Mismatches 33; Indels 10; Gaps 6;
 QY 1061 AGCATGTGAATACCTGAGTCTTGGATATGCGCATATATGCTCAACCCACCGAG 1120
 DB 107 AGATGTGAATACCTGAGTCTTGGATATGCGCATATATGCTCAACCCACCGAG 166
 QY 1121 AGCAATTGTCTACGCTTCTGATGCGTGTGCTCCACACTGTTTAAATGCCATGCTG 1180
 DB 167 AGCAATTGTCTACGCTTCTGATGCGTGTGCTCCACACTGTTTAAATGCCATGCTG 226
 QY 1181 GCATTACTTCTGCTACATAAAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCA 1240
 DB 227 GCATTACTTCTGCTACATAAAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCA 286
 QY 1241 TTGTATCTACAGTATATAGATCGGTACTCAGCAACAAGCTATGCTCTTTTATA 1300
 DB 287 TTGTATCTACAGTATATAGATCGGTACTCAGCAACAAGCTATGCTCTTTTATA 346
 QY 1301 TCAGGTCCCATGATGTGAATATGAGGTGAAGTACTCATCCCAATGATGCCCGGCC 1360
 DB 347 TCAGGTCCCATGATGTGAATATGAGGTGAAGTACTCATCCCAATGATGCCCGGCC 406
 QY 1361 AGTCTCTCTCCCGCCCGCTCATCATGTCAGCGGGTGTCCACAAACAAAGGCTGCCAG 1420
 DB 407 AGTCTCTCTCCCGCCCGCTCATCATGTCAGCGGGTGTCCACAAACAAAGGCTGCCAG 465
 QY 1421 GCTTTATCGGACCAACAGCTTCCCTCTCATCATGATAAAGAAATCCACCTCACTTAATGGGA 1480
 DB 466 GCTTTATCGGACCAACAGCTTCCCTCTCATCATGATAAAGAAATCCACCTCACTTAATGGGA 525
 QY 1481 C-TGGACCATTTGAAAGACAGCCCAAGCAGTTCCATGTCGAGTCCCTAACGGGAATTCAGT 1539
 DB 526 CATGGACCATTTGAAAGACAGCCCAAGCAGTTCCATGTCGAGTCCCTAACGGGAATTCAGT 585
 QY 1540 GTCAACAGGCTAGTCTCTGTTAATGCTTCACTTCTGTCCTCAAACTCGTCAGTTAATAGG 1599
 DB 586 GTCAACAGGCTAG-CTGTAAATGTT-CTTGTCTCAAAATGCTCAGTTAATAGG 640
 QY 1600 TCTCTAGTGTATCCAGACATCTTAAGAAACAAAAAATTACAATCAGTATTCACAAACAG 1659
 DB 641 T-CTCAGTGTATCC-AGAACATCTANGAAACAAAAAATTCAANAGTATNACAACAAGTTG 697
 QY 1660 TTGCC 1664
 DB 698 CTGTC 702
 RESULT 15
 AAH05659
 ID AAH05659 standard; cDNA; 521 BP.
 XX
 AC AAH05659;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (5'-primer) SEQ ID NO:2494.
 XX
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-00116126.
 XX
 PR 29-JUL-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.
 XX
 PS Claim 1; SEQ ID NO 2494; 2537pp + Sequence Listing; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-qt primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602
 CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC	of an oligonucleotide comprising a sequence complementary to the
CC	complementary strand of a polynucleotide which comprises a 5'-end
CC	sequence and an oligonucleotide comprising a sequence complementary to a
CC	polynucleotide which comprises a 3'-end sequence, where the
CC	oligonucleotide comprises at least 15 nucleotides and the combination of
CC	the 5'-end sequence/3'-end sequence is selected from those defined in the
CC	specification. The primer sets can be used in antisense therapy and in
CC	gene therapy. The primers are useful for synthesising polynucleotides,
CC	particularly full-length cDNAs. The primers are also useful for the
CC	detection and/or diagnosis of the abnormality of the proteins encoded by
CC	the full-length cDNAs. The primers allow obtaining of the AHH13628 and
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92445 to AAB95893
CC	represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC	oligonucleotides, all of which are used in the exemplification of the
CC	present invention
XX	
SQ	Sequence 521 BP; 130 A; 135 C; 122 G; 123 T; 0 U; 11 Other;
	Query Match 19.4%; Score 468; DB 4; Length 521;
	Best Local Similarity 96.6%; Pred. No. 8.7e-128;
	Matches 504; Conservative 0; Mismatches 15; Indels 3; Gaps 3;
QY	6 GTGTGCGGGCGGCGCGGCCGAGGGGGATGGAGCGAGCGCGAGCCGGCTCAGAG 65
DB	1 GTGTGCGGCGGCGGCGCGGCCGAGGGGGATGGAGCGAGCGCGAGCCGGTTCAGAG 60
QY	66 TTGACCAATGACCATTGTTGCACAAGCTTCTGAATCTTTCAGACCCATCGACCTATCAGAA 125
DB	61 TTGAACCAATGACCATAGTTTGACAAGCTTCTGAATCTTCAGACCCATCGACCTATCANAA 120
QY	126 TCAGCTCGGCAGCTCCGAGGCAGTCTCACCTGGAGAGCATGATGACGGTTCGCCAGCTG 185
DB	121 TCAGCTCGGCAGCTCCGAGGCAGTCTCACCTGGAGAGCATGATGACGGTTCGCCAGCTG 180
QY	186 GGGTGCTGTGCTTCATTGAATGATGTGTCAAATTCACACACTTCTTTTAGCACCATGACC 245
DB	181 GGGTGCTGTGCTTCATTGAATGATGTGTCAAATTCACACTTCTTTTAGCACCATGACC 240
QY	246 TGGTGCTGTGAGTTTAATTCGAGTTTCATCTGTACTGTACCTGATAAATCAAACCATCACAAAA 305
DB	241 TGGTGCTGTGAGTTTAATTCGAGTTTCATCTGTACTGTACCTGATAAATCAAACCATCACAAAA 300
QY	306 GGATCAAGCCCTAGGTGATGGCATCGCTCTCCACAGAAGTCTTTTTCCCCTATCGAAA 365
DB	301 GGATCAAGCCCTANGTGATGGCATCGCTCTCCACAGAAGTCTTTTTCCCCTATCGAAA 360
QY	366 GAATTTGCTTAAGTGGCAACAACTCATAGAGTGTGGAGCTGGGGCTCCAGAAATTTGGGCAA 425
DB	361 GAATNTGCTTAAGTGGCAACAACTCATNGAGTTGGAGCTGGGGCTCCAGAAATTTGGGCAA 420
QY	426 TAACCTGTTTTGCCAATGCGACACTGCAGTGGTTTTAACCTACACACA-CCTCTTGGCAATT 484
DB	421 TACCTGTTTTGCCAATGCA-CACNCGCANTGTTTTAACCTACACACACCCCTCTTGGCAATT 479
QY	485 ACATGCTATCAATGAACACTC-CAAAACATGTCATGCGAAA 525
DB	480 ACTGCGNATCAATGAACNCCNCNCAAAATATGTCCTGCAAA 521

Search completed: August 16, 2004, 13:39:37
Job time : 1360 secs

Matches	9947	Conserved	ACT
QY	513	ATGTCATGCAGAAAGCCTTTTGATGATGTGTACAAATGCAAGCACATATTACCCAGGCAC	573
Db	17	ATGTCATGCAGAAAGCCTTTTGATGATGTGTACAAATGCAAGCACATATTACCCAGGCAC	76
QY	573	CAGTAATCCTGGGAGCGTTTATAACCAATGTTTGTCATCAATGAGATGGCGCGTATAGC	633
Db	77	CAGTAATCCTGGGAGCGTTTATAACCAATGTTTGTCATCAATGAGATGGCGCGTATAGC	13
QY	633	TAGGCACCTCCGTTTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTGGA	69
Db	137	TAGGCACCTCCGTTTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTGGA	19
QY	693	TGCTATGCAAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCAC	75

RESULT 4	
BX331803	
LOCUS	1144 bp mRNA linear EST 02-MAY-2003
DEFINITION	BX331803 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
	cDNA clone CS0DC002VH20 5-PRIME, mRNA sequence.
ACCESSION	BX331803
VERSION	BX331803.1 GI:30339041
KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE	1 (bases 1 to 1144)

RX331803

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANI

100

REFERENCE

AUTHORS		Li,W.B., Gruber,C., Jessee,J. and Polayes,D.	
TITLE		Full-length cDNA libraries and normalization	
JOURNAL		Unpublished (2001)	
COMMENT		Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4596.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CS0DC002DD100Q1&cluster=4596.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DC002DD100Q1.	
FEATURES		Location/Qualifiers	
source		1..1144	
ORIGIN		/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CS0DC002YH20" /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED" /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo (dfr) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."	
Query Match		37.7%; Score 910.2; DB 13; Length 1144;	
Best Local Similarity		97.7%; Pred. No. 4.3e-233;	
Matches		927; Conservative 15; Mismatches 5; Indels 2; Gaps 2;	
QY	691	GATGCTATGCAGAAAGCATGCTCGAATGGCAGCAATAAATTAGACACACACCCAGGCC	750
DB	64	GATGCTATGCAGAAAGCATGCTCGAATGGCAGCAATAAATTAGACACACACCCAGGCC	123
QY	751	ACCACTCTTGTTGTCAGATATTTGGAGGATACCTTAAGATCTAGATCTAGATCTAAATGTTAAAT	810
DB	124	ACCACTCTTGTTGTCAGATATTTGGAGGATACCTTAAGATCTAGATCTAGATCTAAATGTTAAAT	183
QY	811	TGCAAGGCGTTTCAGATACCTTTTGATCCATATCTTGATATTAACATTTGGAGATAAAGGCT	870
DB	184	TGCAAGGCGTTTCAGATACCTTTTGATCCATATCTTGATATTAACATTTGGAGATAAAGGCT	243
QY	871	GCTCAGAGTGTCAACAGGCAATGGACAGTTTGTGAAGCGGAAACAGCTTGATGGAGAA	930
DB	244	GCTCAGAGTGTCAACAGGCAATGGACAGTTTGTGAAGCGGAAACAGCTTGATGGAGAA	303
QY	931	AATCGGTACAAGTGCAGCAAGTGTAAAGATGGTTCCAGCTTCAAGAGGTTCACTATC	990
DB	304	AATCGGTACAAGTGCAGCAAGTGTAAAGATGGTTCCAGCTTCAAGAGGTTCACTATC	363
QY	991	CATAGATCTCTAATGTTCTTACACTTTCTTGAAACGTTTTCGAAATTTACCGGTGA	1050
DB	364	CATAGATCTCTAATGTTCTTACACTTTCTTGAAACGTTTTCGAAATTTACCGGTGA	423
QY	1051	AAAATTGCTAAGCATGTGAAATACCTCGATATCTTGATATTCGGCCATATGTCCTCA	1110
DB	424	AAAATTGCTAAGCATGTGAAATACCTCGATATCTTGATATTCGGCCATATGTCCTCA	483
QY	1111	CCCAACGGAGAGCAATTTGTCTACGTTCTTGATGTCAGTCTCGTCCACACGTTTAAAT	1170
DB	484	CCCAACGGAGAGCAATTTGTCTACGTTCTTGATGTCAGTCTCGTCCACACGTTTAAAT	543
QY	1171	TGCCATGCTGGCCATTAATCTTCTGCTACATAAAGCTAGCAATGGCTCTGGTATCAATG	1230
DB	544	TGCCATGCTGGCCATTAATCTTCTGCTACATAAAGCTAGCAATGGCTCTGGTATCAATG	603
QY	1231	AATGATCCATTTGATCTACCAAGTATATAGATCGGTACTAGCCAAACAGCCCTATGTG	1290
DB	604	ARTGACTCCATTTGATCTACCAAGTATATAGATCGGTACTAGCCAAACAGCCCTATGTG	663

QY 557 ATATTACCCAGGCACTCAGTAATCCTGGGACGTTATTAAACCAATCTTTGTCTATCAATG 616
 Db 600 ATATTACCCAGGCACTCAGTAATCCTGGGACGTTATTAAACCAATCTTTGTCTATCAATG 659
 QY 617 AGATGGGGCTTAGCTAGTAGGACCTCCGTTTGGAAACCAAGAAGATGCCCATGAATTC 676
 Db 660 AGATGGGGCTTAGCTAGTAGGACCTCCGTTTGGAAACCAAGAAGATGCCCATGAATTC 719
 QY 677 TTCAATACACTGTTGATGCTATCAGAAAGCATGCTGTAATGGAGCAATTAATATAGACA 736
 Db 720 TTCAATACACTGTTGATGCTATCAGAAAGCATGCTGTAATGGAGCAATTAATATAGACA 779
 QY 737 GACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGAG 796
 Db 780 GACACACCCAGGCCACCACTCTTGTGTGATATTTGGAGGATACCTAAGATCTAGAG 839
 QY 797 TCAATGTTTAATTCAGAGGGGCTTTCAGATCTTTGATCCATATCTTTGATATAACAT 856
 Db 840 TCAATGTTTAATTCAGAGGGGCTTTCAGATCTTTGATCCATATCTTTGATATAACAT 899
 QY 857 TGGAGATAAAGGCTGCTCAGAGTGTCACCAAGGCATTGGAGCAAGTTGTGAAG 909
 Db 900 TGGAGATAAAGGCTGCTCAGAGTGTCACCAAGGCATTGGAGCAAGTTGTGAAG 952

RESULT 7
 LOCUS BX451363 918 bp mRNA linear EST 22-MAY-2003
 DEFINITION BX451363 Homo sapiens NEUROBLASTOMA Homo sapiens cdna clone
 CLOBB028ZF10 5-PRIME, mRNA sequence.
 ACCESSION BX451363
 VERSION BX451363.1 GI:31034264
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 JOURNAL Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 COMMENT Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 5254.r For
 more information about this cluster, see
 http://www.genoscope.cns.fr/
 cgi-bin/cluster.cgi?seq=CS0BAF0012F08_AF00063_1&cluster=5254.r.
 Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Paraday Avenue Genoscope sequence ID : CS0BAF0012F08_AF00063_1.
 Location/Qualifiers
 1. 918

FEATURES
 source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CLOBB028ZF10"
 /tissue_type="NEUROBLASTOMA"
 /clone_lib="Homo sapiens NEUROBLASTOMA"
 /notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed
 with a NotI-oligo (dT) primer. Five prime end enriched,
 double-strand cDNA was digested with Not I and cloned into
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.
 Library was not normalized."

ORIGIN
 Query Match 35.3%; Score 850.4; DB 13; Length 918;
 Best Local Similarity 97.9%; Pred. No. 4.7e-217;
 Matches 868; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

QY 864 AAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGACAGCTTGA 923
 Db 21 AAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAAGCGGACAGCTTGA 80
 QY 924 TGGAGAAAACCTCGTACAGTGCAGAAAGTGTAAAGAGATGGTCCAGCTTCAAGAGGTT 983
 Db 81 TGGAGAAAACCTCGTACAGTGCAGAAAGTGTAAAGAGATGGTCCAGCTTCAAGAGGTT 140
 QY 984 CACTATCCATAGATCCTCTAATGTTCTTACACATTTCTCTGAAAAGCTTTTGCAAATTTTAC 1043
 Db 141 CACTATCCATAGATCCTCTAATGTTCTTACACATTTCTCTGAAAAGCTTTTGCAAATTTTAC 200
 QY 1044 CGGTGAAAATTTGCTPAAGAGTGTAAATACCCCTGAGTATCTTGAATTCGGCCATATAT 1103
 Db 201 CGGTGAAAATTTGCTPAAGAGTGTAAATACCCCTGAGTATCTTGAATTCGGCCATATAT 260
 QY 1104 GTCTCAACCAACCGAGAGGCCAATGCTACAGTCTTGTATGTCAGTGTGTCACACTGG 1163
 Db 261 GTCTCAACCAACCGAGAGGCCAATGCTACAGTCTTGTATGTCAGTGTGTCACACTGG 320
 QY 1164 TTTTAAATTCGCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTA 1223
 Db 321 TTTTAAATTCGCATGCTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTA 380
 QY 1224 TCAATGAATGACTCCATTGTATCTACCAAGTGTATAGATCGGTACTCAGCCCAACAAG 1283
 Db 381 TCAATGAATGACTCCATTGTATCTACCAAGTGTATAGATCGGTACTCAGCCCAACAAG 440
 QY 1284 CTATGCTGCTTTTATATACAGTCCCATGATGTGAAAATGAGAGTGAATTTACTCATCC 1343
 Db 441 CTATGCTGCTTTTATATACAGTCCCATGATGTGAAAATGAGAGTGAATTTACTCATCC 500
 QY 1344 CACCCATAGCCCCGCGCAGTCTCTCCCGCCCGCTCATCATGTCAGCGGTTGTCAACAA 1403
 Db 501 CACCCATAGCCCCGCGCAGTCTCTCCCGCCCGCTCATCATGTCAGCGGTTGTCAACAA 560
 QY 1404 CAAACAGCTCGCCAGGCTTTTATCGGACCACAGCTTCCCTCTCATATGATAAAGAAATCC 1463
 Db 561 CAAACAGCTCGCCAGGCTTTTATCGGACCACAGCTTCCCTCTCATATGATAAAGAAATCC 620
 QY 1464 ACCTCACTTAATGGGACTGACCATTTGAAAGACACGCAAGCAGTTCATGTGAGTCC 1523
 Db 621 ACCTCACTTAATGGGACTGACCATTTGAAAGACACGCAAGCAGTTCATGTGAGTCC 680
 QY 1524 TAACGGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGCTTCAGCTTCTGTCCTCAAAA 1583
 Db 681 TAACGGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGCTTCAGCTTCTGTCCTCAAAA 740
 QY 1584 CTGTCAGTTAATAGGTCCTCAGTGTATCCCAAGAACATCTTAAGAAACAAAAAATTTACAA 1643
 Db 741 CTGTCAGTTAATAGGTCCTCAGTGTATCCCAAGAACATCTTAAGAAACAAAAAATTTAC 799
 QY 1644 CAGTATTCACAAAGTTGCTGTTCCGAGTGTCTCAACCTAACCTTCTCATAGTAA 1703
 Db 800 CAGTATTCACAAAGTTGCTGTTCCGAGTGTCTCAACCTAACCTTCTCATAGTAA 859
 QY 1704 TTCTTTGGAGAACCTTACCAAGCCGCTTCCCTCTTCTTACCATTTACCA 1750
 Db 860 NTCTTTGGAGAACCTTACCAAGCCGCTTCCCTCTTCTTACCATTTACCA 906

RESULT 8
 LOCUS BU173371
 DEFINITION AGENCOURT_7984636 NIH MGC_71 Homo sapiens cdna clone IMAGE:6174750
 5', mRNA sequence.
 ACCESSION BU173371
 VERSION BU173371.1 GI:22687355
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 894)	NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		Contact: Robert Strausberg, Ph.D. Email: csapbs-r@mail.nih.gov Tissue procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAMI3548 row: g column: 07 High quality sequence stop: 670.
FEATURES	source			
	1..894	Location/Qualifiers		
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/clone="IMAGE:6174750"			
	/tissue_type="leiomyosarcoma"			
	/lab_host="DH10B (phage-resistant)"			
	/clone_lib="NIH_MGC 71"			
	/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dr. Average insert size 2.1 kb."			
ORIGIN				
Query Match	34.7%;	Score 836.4;	DB 13;	Length 894;
Best Local Similarity	98.9%;	Pred. No. 2.7e-213;		
Matches 884;	Conservative 0;	Mismatches 6;	Indels 4;	Gaps 4;
QY 109	CCATCAGCCTATCAGAAATCAGCTGGAGCTCCGAGGAGTCTCACCCTGGAGACATGGAT	168		
Db 1	CCATCAGCCTATCAGAAATCAGCTGGAGCTCCGAGGAGTCTCACCCTGGAGACATGGAT	60		
QY 169	GCAGTTCTGCCAGCTGGGGTCTGTGTCTTCATTGAATGATGTCAAATCACACACTT	228		
Db 61	GCAGTTCTGCCAGCTGGGGTCTGTGTCTTCATTGAATGATGTCAAATCACACACTT	120		
QY 229	TCCTTTAGGACCACTACCTGGTCTCTAGTTTATTCGAGTTCACTGTACCTGATAAATCA	288		
Db 121	TCCTTTAGGACCACTACCTGGTCTCTAGTTTATTCGAGTTCACTGTACCTGATAAATCA	180		
QY 289	AAACCATCACCAAAAGGATCAAGCCCTAGTGATGCATCGCTCCACAGAAAGTT	348		
Db 181	AAACCATCACCAAAAGGATCAAGCCCTAGTGATGCATCGCTCCACAGAAAGTT	240		
QY 349	CTTTTCCCATCTGAGAAAGATTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGG	408		
Db 241	CTTTTCCCATCTGAGAAAGATTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGG	300		
QY 409	CTCCAGAAATTGGCCAAATCCTGTTTGGCAATGCAGACATCGATGTTTAACTTACACA	468		
Db 301	CTCCAGAAATTGGCCAAATCCTGTTTGGCAATGCAGACATCGATGTTTAACTTACACA	360		
QY 469	CCACCTCTGCCAATTACATGCTATCACATGAACCTCCAAACATGTCATCGAAGGC	528		
Db 361	CCACCTCTGCCAATTACATGCTATCACATGAACCTCCAAACATGTCATCGAAGGC	420		
QY 529	TTTGTGATGATGTCAATGTCAAGCACATATTACCCAGGCACCTCAGTAACTCTGGGAC	588		
Db 421	TTTGTGATGATGTCAATGTCAAGCACATATTACCCAGGCACCTCAGTAACTCTGGGAC	480		
QY 589	GTTATTAAACCAATGTTGTATCAATAGATGGCGGATAGCTAGGCACCTCCGTTTT	648		
Db 481	GTTATTAAACCAATGTTGTATCAATAGATGGCGGATAGCTAGGCACCTCCGTTTT	540		
QY 649	GGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTTGATGCTATGCAGAAAGCA	708		
Db 541	GGAAACCAAGAGATGCCATGAATTCCTTCAATACACTGTTGATGCTATGCAGAAAGCA	600		

QY	1835	GCTCCAGCCGCTGATGAATGCAATCAAGCTGAACTCAGCGTGTCTGGT-GCCCTAT	1893	Db	121	CATGATCAGGTTCTGCGAGCTGGGCTGCTGCTTCATTAAGTGAATGATGTCATCA	180
Db	717	GCKCCAGCCGCTGATGAATGCAATCAAGCTGAACTCAGCGTGTCTGGTGGCCCTAT	776	QY	222	CACACTTTCTTTAGAACAGTACCTGGTGTGTAGTTATTCGAGTTTCATCTGTACCTGA	281
QY	1894	GGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGCAAGAGAT-GG	1952	Db	181	CACACTTTCTTTAGAACAGTACCTGGTGTGTAGTTATTCGAGTTTCATCTGTACCTGA	240
Db	777	GGCGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGCAAGAGATGGG	836	QY	282	TAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGTGTGGCATCGCTCTCCCA	341
QY	1953	GATTGTCAGATGTCAGTCCACTCTCCCGCCCAAGATGCCGAGATGAGGAGGCCAC	2012	Db	241	TAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGTGTGGCATCGCTCTCCCA	300
Db	837	GATTGTCAGATGTCAGTCCACTCTCCCGCCCAAGATGCCGAGATGAGGAGGCCA-	895	QY	342	GAAAGTCTTTTCCCATCTGAGAGATTTTCTTAAGTGGCAACAAACTCATAGAGTTGG	401
QY	2013	TCCGACGAGCTTCAAGAACCCATGACCTCAACCGTGTCTAATAGT	2058	Db	301	GAAAGTCTTTTCCCATCTGAGAGATTTTCTTAAGTGGCAACAAACTCATAGAGTTGG	360
Db	896	TDCGACGAGCTTCAAGAACCCATGACCTCAACCGTGTCTAATAGT	941	QY	402	AGCTGGCTCCAGATTTTGGCAATACCTGTTTGGCAATGAGCACTCGAGTTTAAC	461
RESULT 11				Db	361	AGCTGGCTCCAGATTTTGGCAATACCTGTTTGGCAATGAGCACTCGAGTTTAAC	420
LOCUS	BUI195574	880 bp	mRNA	QY	462	CTACACACACCTCTTTGCCAATTACATGCTATCAATGAGATGCGGCTATAGCTAGGCACCT	521
DEFINITION	AGENCOURT_7967650 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:6171493		linear	Db	421	CTACACACACCTCTTTGCCAATTACATGCTATCAATGAGATGCGGCTATAGCTAGGCACCT	480
ACCESSION	BUI195574			QY	522	AGAAAGCTTTTGTATGATGTACATGTCATGCAAGCACAATATTACCCAGGCACCTCAGTAATCC	581
VERSION	BUI195574.1	GI:22709558		Db	481	AGAAAGCTTTTGTATGATGTACATGTCATGCAAGCACAATATTACCCAGGCACCTCAGTAATCC	540
KEYWORDS	EST.			QY	582	TGGGGAGCTTTATTAAACCAATGTTTGTCAATGAGATGCGGCTATAGCTAGGCACCT	641
SOURCE	Homo sapiens (human)			Db	541	TGGGGAGCTTTATTAAACCAATGTTTGTCAATGAGATGCGGCTATAGCTAGGCACCT	600
ORGANISM	Homo sapiens			QY	642	CCGTTTGGAAACCAAGAGATGCCAATGAAATTCCTTCAATACACTGTTGATGCTATGCA	701
REFERENCE	1 (bases 1 to 880)			Db	601	CCGTTTGGAAACCAAGAGATGCCAATGAAATTCCTTCAATACACTGTTGATGCTATGCA	660
AUTHORS	NIH-MGC http://mhc.nci.nih.gov/.			QY	702	GAAAGCATGCTTGAATGGCAGCAATTAATTAGACAGACACACCCAGGCACCACTCTTGT	761
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)			Db	661	GAAAGCATGCTTGAATGGCAGCAATTAATTAGACAGACACACCCAGGCACCACTCTTGT	720
JOURNAL	Unpublished (1999)			QY	762	TTGTCAGATATTTGGAGTACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	821
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLNL3539 row: o column: 14 High quality sequence stop: 710.			Db	721	TTGTCAGATATTTGGAGTACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	780
FEATURES	Location/Qualifiers			QY	822	TTGTCAGATATTTGGAGTACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	880
source	1..880			Db	781	TTGTCAGATATTTGGAGTACCTTAAGATCTAGAGTCAAAATGTTTAAATTCGAGGGCGT	840
ORIGIN	/organism="Homo sapiens"			RESULT 12			
	/mol_type="mRNA"			AUI124710			
	/db_xref="taxon:9606"			LOCUS	AUI124710	NT2RM4	Homo sapiens cDNA clone NT2RM4000431 5', mRNA
	/clone="IMAGE:6171493"			DEFINITION	AUI124710	822 bp	linear
	/tissue_type="retinoblastoma"			ACCESSION	AUI124710.1	GI:10949426	sequence.
	/lab_host="DH10B (phage-resistant)"			VERSION	AUI124710		EST.
	/clone_lib="NIH_MGC_67"			KEYWORDS	Homo sapiens (human)		
	/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Library constructed by Life Technologies."			SOURCE	Homo sapiens		
				ORGANISM	Homo sapiens		
				REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
				AUTHORS	Ota, T., Yamamoto, J., Nakamura, Y., Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and Isogai, T.		
				TITLE	HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S., Isogai, T.)		
				JOURNAL	Unpublished (2000)		
				COMMENT	Contact: Takao Isogai Genomics Laboratory Helix Research Institute 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan Tel: 81-438-52-3975		
QY	42	ACGAGCCGCGGCTGAGGTTGAAATGACCATGATGTTGACAAAGCTTCTGAATC	101	Db	1	ACGAGCCGCGGCTGAGGTTGAAATGACCATGATGTTGACAAAGCTTCTGAATC	60
QY	102	TTGAGACCATCAGCCATCAGATCAGCTGCGAGTCCGAGGAGTCTCAGCTGAGA	161	QY	1	TTGAGACCATCAGCCATCAGATCAGCTGCGAGTCCGAGGAGTCTCAGCTGAGA	120
Db	61	TTGAGACCATCAGCCATCAGATCAGCTGCGAGTCCGAGGAGTCTCAGCTGAGA	120	QY	162	CATGGATGACGGTTCTGCCAGCTGGGGTGTGTCTTCATTGATGATGTGTCATCA	221

Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM4000431"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RM4"
/note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

ORIGIN

Query Match 33.0%; Score 795.8; DB 9; Length 822;
Best Local Similarity 98.8%; Pred. No. 2.2e-202;
Matches 811; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 856 TTGGAGTAAAGCGTCTCAGAGTGTCACAAAGGCATTGGAGCAGTTTGTGAAGCCGGA 915
Db |||||
QY 916 CAGCTTGTATGGAGAAACCTGCTACAGTGCACNAAGTGTAAAGAGTGTTCAGCTTCA 975
Db |||||
QY 61 CAGCTTGTATGGAGAAACCTGCTACAGTGCACNAAGTGTAAAGAGTGTTCAGCTTCA 120
Db |||||
QY 976 AAGAGTTCATATCCATAGATTCCTTAATGTCTTACACTTCTCTGAAAGCTTTGCA 1035
Db |||||
QY 121 AAGAGTTCATATCCATAGATTCCTTAATGTCTTACACTTCTCTGAAAGCTTTGCA 180
Db |||||
QY 1036 AATTTTACGGTGGAAAAATTGCTAAGAGTGTAAATACCCCTGAGTATCTTGATATCGG 1095
Db |||||
QY 181 AATTTTACGGTGGAAAAATTGCTAAGAGTGTAAATACCCCTGAGTATCTTGATATCGG 240
Db |||||
QY 1096 CCATATATGCTCAACCAAGGAGAGCAATTGTCTAGCTTTGTATGCGAGTGTGTC 1155
Db |||||
QY 241 CCATATATGCTCAACCAAGGAGAGCAATTGTCTAGCTTTGTATGCGAGTGTGTC 300
Db |||||
QY 1156 CACACTGTTTTAATGTCCATGCTGGCCATTACTTCTGCTACATAAAGCTTAGCAATGGC 1215
Db |||||
QY 301 CACACTGTTTTAATGTCCATGCTGGCCATTACTTCTGCTACATAAAGCTTAGCAATGGC 360
Db |||||
QY 1216 CTCTGGTATCAATGAATGATCCTCATTTGTATCTACAGTGTATTTAGATCGGTACTCAGC 1275
Db |||||
QY 361 CTCTGGTATCAATGAATGATCCTCATTTGTATCTACAGTGTATTTAGATCGGTACTCAGC 420
Db |||||
QY 1276 CAACAAGCCTATGTGCTTTTATATCAGGTCCCATGATGTGAAATATGGAGGTGAACCTT 1335
Db |||||
QY 421 CAACAAGCCTATGTGCTTTTATATCAGGTCCCATGATGTGAAATATGGAGGTGAACCTT 480
Db |||||
QY 1336 ACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGTCATCAGTCAGTCAGCGGTT 1395
Db |||||
QY 481 ACTCATCCCAACCCATAGCCCGGCGAGTCTCTCCCGCCCGTCATCAGTCAGTCAGCGGTT 540
Db |||||
QY 1396 GTCAACCAAGAAACAGGCTGGCGAGGCTTTATCGGACCAAGCTTCCCTCTCATCATATA 1455
Db |||||
QY 541 GTCAACCAAGAAACAGGCTGGCGAGGCTTTATCGGACCAAGCTTCCCTCTCATCATATA 600
Db |||||
QY 1456 AAGAATCCACTCACTTAAATGGAGTGGACCATTTGAAGACACAGCCCAAGCTTCCATG 1515
Db |||||
QY 601 AAGAATCCACTCACTTAAATGGAGTGGACCATTTGAAGACACAGCCCAAGCTTCCATG 660
Db |||||
QY 1516 TCAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGCTTCT 1575
Db |||||
QY 661 TCAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTCTGTTAATGCTTCAGCTTCT 720
Db |||||
QY 1576 GTCCAAACCTGGTCACTTAATAGTCTCTAGTATCCCAAGACATCTTAAGAAACAAAA 1635
Db |||||

Db 721 GTCCAAACCTGGCGAGTTAATAGTCTCTCA-TGATCCAGAACATTTCTAAGAAACCAAAA 779
QY 1636 ATTACATCAGTATTTCACAAAGTTTCCCTGTTCCGCGAGTG 1676
Db |||||
780 ATTTCATCAGTATTTCACAAAGTTTCCCTGTTCCGCCANGG 820

RESULT 13
LOCUS AUI21934 818 bp mRNA linear EST 01-AUG-2002
DEFINITION AUI21934 MAMMAL Homo sapiens cDNA clone MAMMAL001308 5', mRNA
sequence.
ACCESSION AUI21934
VERSION AUI21934.1 GI:10937169
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 818)
AUTHORS Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
Isogai,T.
TITLE HRI human cDNA project
JOURNAL Unpublished (2000)
COMMENT Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1. .818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MAMMAL001308"
/tissue_type="mammary gland"
/clone_lib="MAMMAL"
/note="Vector: pME18SFL3"

ORIGIN

Query Match 31.6%; Score 762; DB 9; Length 818;
Best Local Similarity 98.5%; Pred. No. 2.7e-193;
Matches 809; Conservative 0; Mismatches 8; Indels 4; Gaps 4;

QY 727 AAATTAGACAGACACACCCAGGCCACCACTCTTGTTCAGATATTTGGAGGATACCTA 786
Db 1 AAATTAGACAGACACACCCAGGCCACCACTCTTGTTCAGATATTTGGAGGATACCTA 60
QY 787 AGATCTAGAGTCAAAATGTTTAAATTGCAAGGGCTTTTCAGATCTTTTTCATATCTT 846
Db 61 AGATCTAGAGTCAAAATGTTTAAATTGCAAGGGCTTTTCAGATCTTTTTCATATCTT 120
QY 847 GATATAACATTTGAGATAAAGGCTGTCTACAGTGTCAACAAGGCATTGGAGCAGTTTGTG 906
Db 121 GATATAACATTTGAGATAAAGGCTGTCTACAGTGTCAACAAGGCATTGGAGCAGTTTGTG 180
QY 907 AAGCCGGAACAGCTTGTATGGAGAAACTGTACAGTGTCAAGTGTCAAGTGTAAAAAGATGGTT 966
Db 181 AAGCCGGAACAGCTTGTATGGAGAAACTGTACAGTGTCAAGTGTCAAGTGTAAAAAGATGGTT 239
QY 967 CCAGCTTTCAAAGAGGTTCATATCCATAGATCTCTTAATGTTCTTACACTTCTCTGAAA 1026
Db 240 CCAGCTTTCAAAGAGGTTCATATCCATAGATCTCTTAATGTTCTTACACTTCTCTGAAA 299
QY 1027 CGTTTTCGAAATTTTACCGGTGAAAAATTTGCTAAGGATGTGAAATACCTCGATATCTT 1086
Db |||||

RESULT 15
BM563326
LOCUS
LOCUS
DEFINITION
BM563326 1088 bp mRNA linear EST 20-FEB-2002
AGENCOURT 6589448 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441601
5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

FEATURES	source	Location/Qualifiers
	1..763	
		/organism="Homo sapiens"
		/mol_type="mrna"
		/db_xref="taxon:9606"
		/clone="NT2RP3001577"
		/cell_type="teratocarcinoma"
		/cell_line="NT2"
		/clone_lib="NT2RP3"

REFERENCE

1 (bases 1 to 1088)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-k@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 cDNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
 Plate: LLCM1916 row: c column: 10
 High quality sequence stop: 611.

FEATURES

Location/Qualifiers
 1..1088
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5441601"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC 98"
 /note="Organ: Brain; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 29.4%; Score 709.6; DB 12; Length 1088;
 Best Local Similarity 99.3%; Pred. No. 4e-179;
 Matches 723; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 133 GGCAGCTCCGAGGCGAGTCTCAGCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGCT 192
 Db 1 GGCAGCTCCGAGGCGAGTCTCAGCTGGAGACATGGATGAGGTTCTGCCAGCTGGGGTGCT 60

QY 193 GTGCTTCATTGAATGATGTGCAATCACAACACTTTCTTTAGGACCAGTACCTGGTGCT 252
 Db 61 GTGCTTCATTGAATGATGTGCAATCACAACACTTTCTTTAGGACCAGTACCTGGTGCT 120

QY 253 GTAGTTTATTCGAGTTCATCTGTACTGATTAATCAAAACCATCACCAAAAGGATCAA 312
 Db 121 GTAGTTTATTCGAGTTCATCTGTACTGATTAATCAAAACCATCACCAAAAGGATCAA 180

QY 313 GCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGT 372
 Db 181 GCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGT 240

QY 373 CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 432
 Db 241 CTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 300

QY 433 TTTCGCAATGCAGCACTGCAGTGTAACTTACCTACACACACCTCTTGGCAATTAATGCTA 492
 Db 301 TTTCGCAATGCAGCACTGCAGTGTAACTTACCTACACACACCTCTTGGCAATTAATGCTA 360

QY 493 TCACATGAACACTCCAAACATGTGTCAGAGGCTTTTGTATGATGTGTACAAATGCAA 552
 Db 361 TCACATGAACACTCCAAACATGTGTCAGAGGCTTTTGTATGATGTGTACAAATGCAA 420

QY 553 GCACATATTCCAGGCACTCAGTAATCTTGGGACGTTATTAACCAATCTTTGTTCATC 612
 Db 421 GCACATATTCCAGGCACTCAGTAATCTTGGGACGTTATTAACCAATCTTTGTTCATC 480

QY 613 AATCAGATGGCGGTATAGCTAGGCACTCGCTTTTGGAAACCAAGAGATGCCATGAA 672
 Db 481 AATCAGATGGCGGTATAGCTAGGCACTCGCTTTTGGAAACCAAGAGATGCCATGAA 540

QY 673 TTCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAATTA 732
 Db 541 TTCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAATTA 600

QY 733 GACAGACACACCCAGGCCACACACTCTTTGTTGTCCAGATATTTGGAGGATACCTTAAGATCT 792
 Db 601 GACAGACACACCCAGGCCACACACTCTTTGTTGTCCAGATATTTGGAGGATACCTTAAGATCT 660

QY 793 AGAGTCAAAATGTTTAA-TTGCAAGGCGCTTTCAGATACCTTTGATCCATATCTTGATAT 851
 Db 661 AGAGTCAAAATGTTTAAATTTGCAAGGCGCTTTCAGATACCTTTGATCCATATCTTGATAT 720

QY 852 AACATTGG 859
 Db 721 AACATTGG 728

Search completed: August 16, 2004, 20:20:44
 Job time : 9145 secs

THIS PAGE BLANK (USPTO)

407 AGTCCTCTCCCGCCCGCTCATCACTCAGCGGTTGTCCACCAACAAACAGGCTGGC-CAG 465
 1421 GCTTTATCGGACACAGCTTCCCTCTCATCATATAAGAAATCCACCTCACTTAATGGGA 1480
 466 GCTTTATCGGACACAGCTTCCCTCTCATCATATAAGAAATCCACCTCACTTAATGGGA 525
 1481 C-TGGACCAATTCGAAGACACAGCGCAAGCAGTTCATGTGAGTCCCTTAACGGGAATTCAGT 1539
 526 CATGACCAATTCGAAGACACAGCGCAAGCAGTTCATGTGAGTCCCTTAACGGGAATTCAGT 585
 1540 GTCAACAGGGCTAGTCTGTTAATGTTTCTGCTTCTGCTCAAAATCGTCAAGTATATAGG 1599
 586 GTCAACAGGGCTAGTCTGTTAATGTTTCTGCTTCTGCTCAAAATCGTCAAGTATATAGG 640
 1600 TCTCAGTATCCAGAACATCTTAAGAAACAAACAAATTAACATCAGTATTCACAAACAG 1659
 641 T-CTCAGTATCC--AGAATCTTANGAAACAAACAAATTAACATCAGTATTCACAAACAG 697
 1660 TTGCC 1664
 698 CTGTC 702

RESULT 2

US-09-408-865-2
 ; Sequence 2, Application US/09408865A
 ; Patent No. 6329171
 ; GENERAL INFORMATION:
 ; APPLICANT: Kapeller-Libermann, Rosana
 ; TITLE OF INVENTION: 23484, A NOVEL HUMAN UBIQUITIN
 ; FILE REFERENCE: 5800-51
 ; CURRENT APPLICATION NUMBER: US/09/408.865A
 ; CURRENT FILING DATE: 1999-09-30
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 3941
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; NAME/KEY: CDS
 ; LOCATION: (279)...(3650)
 ; US-09-408-865-2

Query Match 17.5%; Score 421.8; DB 4; Length 3941;
 Best Local Similarity 63.8%; Pred. No. 7.8e-123;
 Matches 639; Conservative 0; Mismatches 362; Indels 0; Gaps 0;
 QY 301 CAAAGGATCAAGCCCTAGTGATGGCATCGCTCTCCACAGAAAGTCTTTCCCATCT 360
 DB 540 CACAGTATGAGAGCTGTGTGACGAGTCCAGCCCGCCAGAAAGTGTCTTTCCCGCAG 599
 QY 361 GAGAGATTTGTTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTTG 420
 DB 600 GAGCGACTCTCTGAGGTGGAGCGGGTCTCCGCGTGGCGGAGTCTCCACACCTT 659
 QY 421 GCGAATACCTGTTTCCCAATGACGACCTGAGTGTAACTTACCTACACACCTCTTGGC 480
 DB 660 GGCACACCTGCTTCTCAATGACCACTCAGTGTGACCTACACACCTCTTAGCC 719
 QY 481 AATTATCATGCTATCATGACAACTCCAAACATGTATGACAGAGGCTTTGTATGATG 540
 DB 720 AACTACCTGCTCTCCAGGAGCATGTCTGCGAGCTGCCACAGGGAAGCTTCTCATGCTG 779
 QY 541 TGTACAAATCAAGCAGCATATTTACCGGCACTCAGTAATCTCGGGGACGTTTAAACCA 600
 DB 780 TGTGATGACGAAACCAATTTCCAGCCCTTCGCCAAGACGCGCAAGCCATCAAGCCC 839
 QY 601 ATGTTTGTCAATAGATGCGCGGTATAGTACGACCTCGTGTGTTGAAACCAAGAA 660
 DB 840 GTCTCTTCATCCGAGACCTGAAAGAGATGCGCGACACTTCCGCTTTGGGAACCGAG 899
 QY 661 GATGCCCATGAATTCCTTCAATACACTGTTCATGCTATGCAGAAAGCATGCTTGAATGGC 720

900 CACGCGCATGAGTTCTCGGTACACCATCGACCGCATGACAGAAACCTGCTGCTGATGCG 959
 721 AGCAATAATATAGACAGACACACCCAGGCGCCACTCTTTGTTGTGAGATATTTGGAGGA 780
 960 TGTGCAAGTGGATGCTGTCAAACGAGGCTACTACCTTGGTCCATCAAAATTTTGGAGG 1019
 781 TACCTAAGATCTAGAGTCAAAATGTTTAAATTCGAAGGGGCTTTCAGATACCTTTGATCCA 840
 1020 TATCTCAGATCACGCGTGAAGTCTCCGTTGTCAGAGGCTCTCGGACACCTACGACCCC 1079
 841 TATCTTGAATATAACATTGAGATAAAGGCTGCTCAGAGTGTCAAACAGCAATTTGGAGAG 900
 1080 TACTTGGAGCTGCGCTGGAGATCCGGCAAGCTGCGCAATATTGTGCTGCTGTHACTT 1139
 901 TTTGTCAAGCCGACAGCTTGTATGAGAGAAATCTGTACAGTGCAGAGTGTAAAGAG 960
 1140 TTTGTGAAGAGCAGATGCTCTGAGTGGAGAGATGCCCTACATGTGTGCTAAATGCAAGAG 1199
 961 ATGTTTCCAGCTTCAAAGAGGTTCACTATCATAGATCCCTCTTAATGTTCTTACACTTCT 1020
 1200 AAGTTTCCAGCGCAAGCGCTTCAACATCCACAGAAATCCCAACGCTTAAACCTTTCC 1259
 1021 CTGAACGTTTTCGAATTTTACCGGTGGAATAATTCCTAAGGATGTGAATAATCCCTGAG 1080
 1260 CTCGAAGCGCTTTCGAATTTTACCGGTGGAATAATTCCTAAGGATGTGAATAATCCCTGAG 1319
 1081 TATCTTGAATATTCGCGCATATATGTCTCAACCCACGAGAGCAATGTCTAGCTCTTG 1140
 1320 TTCTCTCAACATACATCTCGTATATGTCCAGAAATATGCTGATCTCTGATATGAGCTC 1379
 1141 TATGAGTGTGTCACACCTGTTTAAATGCGCATGCTGCGCAATCTCTGCTACATA 1200
 1380 TATGCTGTCTGTCACCTCGGCTACAGCTGCGCATGCGGGGCACTATTAATGCTAGCTG 1439
 1201 AAAGCTAGCATGCTCTGCTGATCAATGATGATGATGATGATGATGATGATGATGATGAT 1260
 1440 AAGGCAAGCAATGACAGTGTGACAGATGATGATGATGATGATGATGATGATGATGATGAT 1499
 1261 AGATCGGTACTCAGCAACAGCCTATGCTGCTCTTTTATAT 1301
 1500 AAGTGTGTTCTGACACAGCAGCGCTTACGCTGCTTCTATCT 1540

RESULT 3

US-09-280-116-214
 ; Sequence 214, Application US/09280116A
 ; Patent No. 6331427
 ; GENERAL INFORMATION:
 ; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
 ; FILE REFERENCE: 5800-24, 035800/176965
 ; CURRENT APPLICATION NUMBER: US/09/280,116A
 ; CURRENT FILING DATE: 1999-03-26
 ; NUMBER OF SEQ ID NOS: 268
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 214
 ; LENGTH: 457
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
 ; US-09-280-116-214

Query Match 16.3%; Score 393.8; DB 4; Length 457;
 Best Local Similarity 37.2%; Pred. No. 1.3e-114;
 Matches 444; Conservative 0; Mismatches 7; Indels 6; Gaps 4;
 QY 38 ATGAGCGAGCGCCGAGCGCGGTGAGATGAAATGACCAATGACCAATGATGACCAAGCTCTG 97
 DB 1 ATGAGCGAGCGCGAGCGCGGTGAGATGAAATGACCAATGACCAATGATGACCAAGCTCTG 60
 QY 98 AATCTTCAGACCCATCAGCCT--ATCAGAAATCAGCCTGCGAGCTCCGAGGCGAGTCTCACC 155

Db 61 AATCTTCAGACCATCAGCCTTATCATGAATCAGCTCGCATCTCCGAGCGAGTCTCACCC 120
 QY 156 TGGAGCATGGATGCGAGTTCTGCCAGCTGGGGTGTGTCTTTCATTAATGATGTGC 215
 Db 121 TGGAGCATGGATGCGAGTTCTGCCAGCTGGGGTGTGTCTTTCATTAATGATGTGC 180
 QY 216 AATCAGACACATCTTTTATAGGACCACTAGCTGGTGTGTGTATTCAGTTCATCTGT 275
 Db 181 AATCAGACACATCTTTTATAGGACCACTAGCTGGTGTGTGTATTCAGTTCATCTGT 240
 QY 276 ACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCC 335
 Db 241 ACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCC 300
 QY 336 TCCACA-GAAAGTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAAAACTCATA 394
 Db 301 TCCACAGGAAAGTACTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAAAACTCATA 360
 QY 395 GAGTTGGAGCTGGCTCCAGAA--TTTGGCAATACCT-GTTTTGCCAATGAGCACTGC 451
 Db 361 GAGTTGGAGCTGGCTCCAGAAATTGGGGCAATACCTGGTTTGGCAATGAGCACTGC 420
 QY 452 AGTGTTTAACTTACACACACACCTCTTTGCCAATTACAT 488
 Db 421 AGTGTTTAACTTACACACACCTCTTTGCCAATTACAT 457

RESULT 4

US-09-280-116-219
 ; Sequence 219, Application US/09280116A
 ; Patent No. 6331427

GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.
 ; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
 ; FILE REFERENCE: 5800-24, 035800/176965
 ; CURRENT APPLICATION NUMBER: US/09/280,116A
 ; CURRENT FILING DATE: 1999-03-26
 ; NUMBER OF SEQ ID NOS: 268
 ; SOFTWARE: Patent In Ver. 2.0
 ; SEQ ID NO 219
 ; LENGTH: 917
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
 US-09-280-116-219

Query Match 13.3%; Score 320; DB 4; Length 917;
 Best Local Similarity 64.0%; Pred. No. 5.9e-91;
 Matches 499; Conservative 0; Mismatches 280; Indels 1; Gaps 1;

QY 523 GAAGCTTTTGTATGATGTGATCAATGCAAGCATATTAACCGGCACTCAGTAATCCT 582
 Db 9 GGAAGCTTCTGCATGCTGTGTGTATGATGCAAGCATATTAACCGGCACTCAGTAATCCT 582
 QY 583 GGGAGCGTTATTAACCAATGTTTGTCTCATATGATGCGGGGTATAGTAGGACCTC 642
 Db 69 GGCACGCCATCAAGCCCGTCTCTTCATCGAGACCTGAAAGAGATCGCCGACACTTC 128
 QY 643 CGTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACATCTGTTGATGCTATGCAG 702
 Db 129 CGCTTTGGAAACCAAGAGATGCCATGAATTCCTTCAATACATCTGTTGATGCTATGCAG 188
 QY 703 AAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACCCAGGCCACCACTCTTGT 762
 Db 189 AAAGCCTGCTGAATGGTGTGTCGCAAGTTGATGTCGTCGTCGTCGTCGTCGTCGTC 248
 QY 763 TGTGAGATATTTGGAGGATCTAAGATCTAGAGTCAATGTTTAAATGCAAGGCGTT 822
 Db 249 CATCAATTTTGGAGGATCTAGAGTCAATGTTTAAATGCAAGGCGTT 822
 QY 823 TCAGATACCTTTTGTATCCATATCTTGTATATAACATTTGGAGATAAAGGCTGCTCAGAGTGC 882

Db 309 TCGGACACCTAGACACCCCTACTTGGACATCGGCTGGAGATCGGCAAGCTGGAATATT 368
 QY 883 AACAGGCATTTGGAGCAGTTTCTGAAGCCGGAACAGCTTGTATGGAGAAAACCTCGTACAG 942
 Db 369 GTGCGTGTCTGGAACCTTTTGTGAAAGCAGATGTCTGAGTGGAGAGATGCTTACATG 428
 QY 943 TCGACGAGTGTAAAGAGATGTTCCAG-CTTCAAAGAGTTTCACTATCCATAGATCCTC 1001
 Db 429 TGTGCTAAATGCAAGAAGAGGCTCCAGCCAGCAAGGCTTCAACCATTCACAGAAT 488
 QY 1002 TAAATGTTCTTACACATTTCTCTGAAAGTTTTCGAAATTTTACGGTGGAAAAATTCGCTAA 1061
 Db 489 CAACGCTCTTAACCCCTTTCCCTCAAGCGCTTTGCAACTTCAGCGGGGGAAGATCACCAA 548
 QY 1062 GGATGTGAAATACCTCGATGATCTTGTATATTCGGCCATATATGTCTCAACCAACGGAGA 1121
 Db 549 GGATGTAGGCTATCCGGAATTCCTCAACATAGCTCCGTATATGTCCAGAAATAATGGTGA 608
 QY 1122 GCCAATGTCTACGCTTTGTATGAGTGTGTGTCACACTGTTTAAATTCGCAATGCTGG 1181
 Db 609 TCCTGTATGATGACTCTATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 668
 QY 1182 CCATTACTTCTGTACATATAAAGCTAGCAATGGCTCTGCTATCAATGAATGATCCAT 1241
 Db 669 GCACATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 728
 QY 1242 TGTATCTACAGTGTATTTAGATCGTACTCAGCCAAACAGCTTATGCTCTTTTATAT 1301
 Db 729 GGTCCATTCAGCAACGCTCAAGTGTCTTGAACGAGGCTACGCTGCTGCTGCTGCTGCT 788

RESULT 5

US-09-019-095A-37
 ; Sequence 37, Application US/09019095A
 ; Patent No. 6287858

GENERAL INFORMATION:

; APPLICANT: D'Andrea, Alan D.
 ; APPLICANT: Zhu, Yuan
 ; TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate
 ; FILE REFERENCE: DFCI-435p2A2
 ; CURRENT APPLICATION NUMBER: US/09/019,095A
 ; CURRENT FILING DATE: 1998-02-05
 ; PRIOR APPLICATION NUMBER: PCT/US96/12884
 ; PRIOR FILING DATE: 1996-08-07
 ; PRIOR APPLICATION NUMBER: US 60/002,066
 ; PRIOR FILING DATE: 1995-08-09
 ; PRIOR APPLICATION NUMBER: US 60/019,787
 ; PRIOR FILING DATE: 1996-06-14
 ; NUMBER OF SEQ ID NOS: 51
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 37
 ; LENGTH: 1873
 ; TYPE: DNA
 ; ORGANISM: murine
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (199)....(1833)
 US-09-019-095A-37

Query Match 13.0%; Score 313.6; DB 3; Length 1873;
 Best Local Similarity 56.2%; Pred. No. 1e-88;
 Matches 617; Conservative 0; Mismatches 469; Indels 12; Gaps 1;

QY 232 TTAGGACCACTACCTGGTGTGTATTTATTCGAGTTTCATCTGTACCTGATAAATCAAAA 291
 Db 178 TTTGAAGAGTCTTTGGAGACATGGTGTCTCTTCTTCTCCAGAGCAAGATCCAGCC 237
 QY 292 CCATCACCAAAAAGGATCAAGCCCTAGGTGATGCGATCGCTCCCTCCACAGAAAGTCTT 351
 Db 238 CTATCATCTCTGTGTGCCCAACAGCTGCATCAGATGAAGCTCAGGTAGTGGAGCTA 297


```

; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 217
; LENGTH: 672
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
; NAME/KEY: misc feature
; LOCATION: (1)..(672)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-217

Query Match      8.1%; Score 194.2; DB 4; Length 672;
Best Local Similarity 62.5%; Pred. No. 4.2e-51;
Matches 298; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 301 CAAAGGATCAAGCCCTAGTGTGATGGATGCTCTCCACAGAAAGTTCTTTCCCATCT 360
DB 196 CACACGTATGAGAGCTGTGTGACGGAGTCCACAGCCCGCAGAAAGTGTCTTTCCCAAG 255
QY 361 GAGAAATTTGCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTG 420
DB 256 GAGCGACTGTCTGAGGTGGGAGCGGGTCTTCGGCTGGCGCAGGACTCCCAACCTT 315
QY 421 GCGAATACCTGTGTTGCAATGAGCAGCACTGAGTGTAACTTACATACACACCTCTTGCC 480
DB 316 GCGACACCTGCTTCTCTATGCGCACCATCCAGTGTGACCTTACATACACACCTCTAGCC 375
QY 481 AATTACATGTATCAGATGAACACTCCAAACATGTATGATGATGATGATGATGATGATG 540
DB 376 AACTTACCTGTCTCCAGGAGCATGCTCGCAGCTGCCACAGGAAAGCTTCTGATGCTG 435
QY 541 TGTAACATGCAAGACATATATCCAGGCACTCAGTAACTCTGGGAGCTGTTATTAACCA 600
DB 436 TGTGTATGAGAACCACTATGTCAGGCTTCGCCAACAGGCGCAAGCCATCAAGCCC 495
QY 601 ATGTTTGTCTCAATGAGATGCGGCGTATAGTACAGCCTCGCTTTTGGAAACCAAGAA 660
DB 496 GTCTCTCTTCCGANAACATGAAAAGATGCCCGACACTTCCGCTTTGGGAAACCAAGAA 555
QY 661 GATGCCATGAATTCCTTCAATACACTGTTGATGCTATGAGAAAGCATGCTTGAATGGC 720
DB 556 GAGCGCATGAATTCCTCGCGGTACACCATGACACACCAACCAAGCAAGCAAGTGGC 615
QY 721 AGCAATAAATTAGACAGACACACACCAAGCCACCACTCTTGTGTTGTCAGATATTTGGA 777
DB 616 TGTGCGAAGTTGGATCTTCAACNAGCTACTAATTTGGTCCATCAAAATTTTGGGA 672

RESULT 9
US-09-058-489-33
; Sequence 33, Application US/09058489
; Patent No. 6103886
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; APPLICANT: Lahn, Bruce
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WHI97-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 33
; LENGTH: 10091
; TYPE: DNA
; ORGANISM: Human
; US-09-058-489-34

Query Match      1.9%; Score 44.8; DB 3; Length 10091;
Best Local Similarity 47.5%; Pred. No. 0.0079;
Matches 133; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

QY 857 TGGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTGAGCGGGAAC 916
DB 6895 TGGATATTAGAAATCATCAAAATCTTCTTGACTCTTTGGAAACAGTATATCAAGAGATT 6954
QY 917 AGCTTGATGGAGAAACTCGTACAAAGTGCAGCAAGTGTAAAAAGATGTTCCAGCTTCAA 976
DB 6955 TATTGGAAGGTGCAAAATGCATATCTTTGAAAATGTGATAAAAGGTTGCACAGTAA 7014
QY 977 AGAGTTTCACTATCCATAGATCCTCTAATGTTTCTTACACTTCTCTGAAACGTTTGGAA 1036
DB 7015 AGCGCTCTCTAATTAATAAATTCCTCGGTTCTTGCTATCCAACTCAACAGATTGACT 7074
QY 1037 ATTTTACCGGTGCAAAAATTTGTAAGGATGTGAATACCTCGAGTATCTTGATATTCGGC 1096
DB 7075 ATGACTGGGAAAGAGATGTGCAATTAATTAATTAATTTTGAATTTCTTCGAGAGC 7134
QY 1097 CATATATGCTCAACCCACAGGAGCAATTTGTCTAGCT 1136
DB 7135 TGGATATGGGACCTTACACAGTAGCAGGTGTGCAAACT 7174

RESULT 11
US-09-280-116-222
; Sequence 222, Application US/09280116A
```



```

; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 222
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
; NAME/KEY: misc feature
; LOCATION: (1)-(1998)
; OTHER INFORMATION: n = a, t, c or g
US-09-280-116-222

```

```

Query Match      1.7%; Score 42.2; DB 4; Length 1998;
Best Local Similarity 50.8%; Pred. No. 0.017;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 835 GATCCATATCTTGATATACATGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTG 894
Db 273 GAAGCTTTTCATGCGCTCAATCTAGGAGTGACTTCTGTGTCAGAGTTTGAAATTTCTTG 332
QY 895 GAGCAGTTTGTGAAGCGGAACAGCTTGATGAGAAACTCTCAAGTGCAGCAAGTGT 954
Db 333 GACCAATTTGTAGAGGAGAGTTCTAGAGGAAGTAATGCGTACTCTGTGAAAAGTGT 392
QY 955 AAAAGATGGTTCAGCTTCAAGAGGTTCACTATCCATAGATCTCTAAATCTTCTTACA 1014
Db 393 AAAGAAAAGAGAATAACAGTGAAGGACCTGTATTAATCTTACCTAGCGTCTTGTA 452
QY 1015 CTTTCTCTGAAACGTTTG 1033
Db 453 ATTCACCTAATGAGATTG 471

```

```

RESULT 12
US-09-976-594-862
; Sequence 862, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 862
; LENGTH: 5734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1073168.8
US-09-976-594-862

```

```

Query Match      1.7%; Score 42.2; DB 4; Length 5734;
Best Local Similarity 50.8%; Pred. No. 0.035;
Matches 101; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 835 GATCCATATCTTGATATACATGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTG 894
Db 641 GAAGCTTTCATGCGCTCAATCTAGGAGTGACTTCTGTGTCAGAGTTTGAAATTTCTTG 700

```

```

QY 895 GAGCAGTTTGTGAGCGGAACAGCTTGTATGAGAAAACTCGTACAGTCGAGCAAGTGT 954
Db 701 GACCAATTTGTAGAGGAGAGTTCTAGAGGAAGTAATGCGTACTACTGTGAAAAGTGT 760
QY 955 AAAAGATGGTTCAGCTTCAAGAGGTTCACTATCCATAGATCTCTAAATCTTCTTACA 1014
Db 761 AAAGAAAAGAGAATAACAGTGAAGGACCTGTATTAATCTTACCTAGCGTCTTGTA 820
QY 1015 CTTTCTCTGAAACGTTTG 1033
Db 821 ATTCACCTAATGAGATTG 839

```

```

RESULT 13
US-09-976-594-861
; Sequence 861, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 861
; LENGTH: 1683
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1073168.6
; NAME/KEY: unsure
; LOCATION: 118, 142
; OTHER INFORMATION: a, t, c, g, or other
US-09-976-594-861

```

```

Query Match      1.7%; Score 41.8; DB 4; Length 1683;
Best Local Similarity 53.3%; Pred. No. 0.02;
Matches 88; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 869 CTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGAG 928
Db 158 CTGCTCAGAGTTTGGAAATTTCTTTGGACCAATTTGTAGAGGAGAAAGTTCTAGAAGAA 217
QY 929 AAAACTCGTCAAGTGCAGCAAGTGTAAAAAGATGGTTCCAGCTTCAAGAGGTTCACTA 988
Db 218 GTAATGCGTACTACTGTGAAGAGTGTAAAGAAAAGAGAATAACAGTGAAGAACCTGTGA 277
QY 989 TCCATAGATCTCTAATGTTCTTACACTTTTCTCTGAAACGTTTG 1033
Db 278 TTAATCTTTTACCTAGCGTCTGTGTAATTCACCTAATGAGATTG 322

```

```

RESULT 14
US-08-858-207A-234/c
; Sequence 234, Application US/08858207A
; Patent No. 6348328
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Nicholas, Richard
; APPLICANT: Stodola, Robert
; TITLE OF INVENTION: No. 6348328el Compounds
; NUMBER OF SEQUENCES: 552
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road

```

us-10-049-745-31.rni

Wed Aug 18 13:53:01 2004

```

;
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6492505 79777CT1
US-09-495-050A-40

Query Match      1 6%; Score 38; DB 4; Length 847;
Best Local Similarity 98 0%; Pred. No. 0.2;
Matches 49; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 2363 ATGCGCGCGCAAGCTATTCCCTTTAGTGAGGGTTAATTTAGCTTGA 2412
DB 56 ATGCGCGCGCAAGCTATTCCCTTTAGTGAGGGTTAATTTAGCTTGA 8

Search completed: August 16, 2004, 20:24:14
Job time : 201 secs

```

```

;
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/08/858,207A
; APPLICATION NUMBER: US/08/858,207A
; FILING DATE: 09-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/017670
; FILING DATE: 14-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50475
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 234:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 730 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-858-207A-234

Query Match      1.6%; Score 38.2; DB 4; Length 730;
Best Local Similarity 49.7%; Pred. No. 0.15;
Matches 97; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1020 TCTGAAACGTTTTCGCAAAATTTACCGGTGGGAAATTTGCTAAGGATGTGAAATACCTGA 1079
DB 212 TATGAACGTTATTTGCTTTTGTAGTGGCGGAGACGACCTGGTATGAACGCTGCTAT 153

QY 1080 GTATCTTGATTCGGCCATATATCTCTCAACCCAGGAGAGCAATTTCTACGTCCT 1139
DB 152 CCGTGCAGTTGTTTGGTCAAGCAATTTCAAGAGGAATGGAAGTGTGTTGATCTATGACGG 93

QY 1140 GTATGCAATGCTGCTCCACACTGGTTTAAATTCATGCTGGCCATTAATTTCTGCTACAT 1199
DB 92 ATATGCTGTTAGTTGCTGGGTGAAATTCATCCCTAGATGCAAGTTCAGTAGGAGACAT 33

QY 1200 AAAAGCTAGCAATGG 1214
DB 32 CATTTCTGTTGGTGG 18

RESULT 15
US-09-495-050A-40/c
; Sequence 40, Application US/09495050A
; Patent No. 6492505
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/09/495,050A
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: February 1, 1999
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Homo sapiens

```

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 20:24:22 ; Search time 14895 Seconds
(without alignments)
7018.691 Million cell updates/sec

Title: US-10-049-745-31
Perfect score: 2412
Sequence: 1 ggsgctgtgcgcgcgccg.....agggttaatttagcttgca 2412

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 30 *ov mcr2*
Total number of hits satisfying chosen parameters: 37

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.un.*
- 28: em.vi.*
- 29: em.htg.hum.*
- 30: em.htg.inv.*
- 31: em.htg.other.*
- 32: em.htg.mus.*
- 33: em.htg.pln.*
- 34: em.htg.pln.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2412	100.0	2412	6	AX083128	Sequence
2	2321	96.2	4659	6	AX786887	Sequence
3	2287	94.8	2347	6	AX107851	Sequence
4	2265	93.9	3660	6	AX882263	Sequence
5	2265	93.9	3660	6	BD159625	Primer fo
6	2265	93.9	3660	9	AK022759	Homo sapi
7	2259	93.7	3669	9	BC060846	Homo sapi
8	1475	61.2	1679	6	AX364931	Sequence
9	722	29.9	152927	9	AC004895	Homo sapi
10	722	29.9	154791	2	AC069393	Homo sapi
11	415	17.2	217022	2	AC146146	Pan trogl
12	357	14.8	722	6	AR264045	Sequence
13	253	10.5	217022	2	AC146146	Sequence
14	245	10.2	5269	6	BD181571	Novel gen
15	232	9.6	521	6	AX867589	Sequence
16	232	9.6	521	6	BD147651	Primer fo
17	205	8.5	457	6	AR264036	Sequence
18	205	8.5	154791	2	AC069393	Homo sapi
19	137	5.7	213	6	AX495102	Sequence
20	100	4.1	292	6	AX387329	Sequence
21	38	1.6	195906	2	AC136746	Mus muscu
22	34	1.4	195906	2	AC136746	Mus muscu
23	32	1.3	239456	2	AC108995	Sequence
24	32	1.3	270456	2	AC111575	Sequence
25	30	1.2	637	6	AX327423	Sequence
26	30	1.2	733	6	AX174745	Sequence
27	30	1.2	823	6	AX078266	Sequence
28	30	1.2	972	6	AX135572	Sequence
29	30	1.2	1106	6	AX135583	Sequence
30	30	1.2	1163	6	AX121858	Homo sapi
31	30	1.2	1487	9	AF121858	Sequence
32	30	1.2	1491	6	AX061211	Sequence
33	30	1.2	1702	6	AX073668	Sequence
34	30	1.2	1950	6	AX081430	Sequence
35	30	1.2	1954	6	AX077793	Sequence
36	30	1.2	1987	6	BD260129	Matrix-re
37	30	1.2	3465	6	BD224624	Protein t

ALIGNMENTS

RESULT 1	AX083128	AX083128	2412 bp	DNA	linear	PAT 28-FEB-2001
LOCUS	Sequence 31 from Patent WO0110903.					
DEFINITION	Sequence 31 from Patent WO0110903.					
ACCESSION	AX083128					
VERSION	AX083128.1	GI:13185035				
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	1					
AUTHORS	Yue, H., Lal, P., Tang, Y.T., Bandnan, O., Baughn, M.R., Azimzai, Y., Lu, D.A. and Yang, J.					
TITLE	Proteases and protease inhibitors					
JOURNAL	Patent: WO 0110903-A 31 15-FEB-2001;					
FEATURES	Incyte Genomics, Inc. (US)					
source	Location/Qualifiers					
	1. .2412					
	/organism="Homo sapiens"					
	/mol_type="unassigned DNA"					
	/db_xref="taxon:9606"					
	/note="Incyte ID No: 1299481CB1"					

	Query Match	100.0%;	Score 2412;	DB 6;	Length 2412;	
	Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches	0;	Gaps 0;
	Matches 2412;	Conservative	0;			
QY	1	GGGCTGTGCGGCGCGGCGCGGCGGCGGCGGAGGGGGATGAGCGAGCGCGAGCCGGT	60			
Dd	1	GGGCTGTGCGGCGCGGCGGCGGCGGCGGAGGGGGATGAGCGAGCGCGGAGCGGGT	60			
QY	61	CAGAGTTGAACAATGACCATTAGTTCACAAAGCTTCTGAATCTTCAGAGCCCATCAGCCTAT	120			
Dd	61	CAGAGTTGAACAATGACCATTAGTTCACAAAGCTTCTGAATCTTCAGAGCCCATCAGCCTAT	120			
QY	121	CAGAATCAGCTGGCAGCTCCGAGCAGTCTCACTGGAGACATGGATGAGGTTCTGCC	180			
Dd	121	CAGAATCAGCTGGCAGCTCCGAGCAGTCTCACTGGAGACATGGATGAGGTTCTGCC	180			
QY	181	AGCTGGGTCGTGTCTTCATTGAATCATGTGTCAAATCACACATTCCTTTAGACCA	240			
Dd	181	AGCTGGGTCGTGTCTTCATTGAATCATGTGTCAAATCACACATTCCTTTAGACCA	240			
QY	241	GTACTGTGTGTAGTTTTATTCGAGTTTCCTCTACTGATAAATCAAACCATCACCA	300			
Dd	241	GTACTGTGTGTAGTTTTATTCGAGTTTCCTCTACTGATAAATCAAACCATCACCA	300			
QY	301	CAAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCA GAGAAAGTTCCTTTCCCATCT	360			
Dd	301	CAAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCA GAGAAAGTTCCTTTCCCATCT	360			
QY	361	GAGAAAGTTTCTTAAGTGGCACAAAACTCATAGAGTTGGAGCTGGCTCCAGAATTG	420			
Dd	361	GAGAAAGTTTCTTAAGTGGCACAAAACTCATAGAGTTGGAGCTGGCTCCAGAATTG	420			
QY	421	GGCAATACCTGTTTTGCCAATGACGACTCGAGTGTTTAACTTACHACA CACTCTTGCC	480			
Dd	421	GGCAATACCTGTTTTGCCAATGACGACTCGAGTGTTTAACTTACHACA CACTCTTGCC	480			
QY	481	AATTACATGCTATCACATGAACACTCCAAAACTCATGACAGAGGCTTTTGATGATG	540			
Dd	481	AATTACATGCTATCACATGAACACTCCAAAACTCATGACAGAGGCTTTTGATGATG	540			
QY	541	TGTACAAATGCAAGCACATATTACCGAGGCACTCAGTAATCTCTGGGACGTTATTAAACCA	600			
Dd	541	TGTACAAATGCAAGCACATATTACCGAGGCACTCAGTAATCTCTGGGACGTTATTAAACCA	600			
QY	601	ATGTTTGTCTCAATCAGATGCGGCTGTATAGCTTAGGCACCTCCGTTTGGAAACCAAGAA	660			
Dd	601	ATGTTTGTCTCAATCAGATGCGGCTGTATAGCTTAGGCACCTCCGTTTGGAAACCAAGAA	660			
QY	661	GATGCCCATGAATTCCTTCAATACACTGTGTGATGCTATGCAAAAGCATGCTTGAATGGC	720			
Dd	661	GATGCCCATGAATTCCTTCAATACACTGTGTGATGCTATGCAAAAGCATGCTTGAATGGC	720			
QY	721	AGCAATAAATTAGACAGACACACCCAGGCCACCATCTGTGTTGTCAGATATTGGAGGA	780			
Dd	721	AGCAATAAATTAGACAGACACACCCAGGCCACCATCTGTGTTGTCAGATATTGGAGGA	780			
QY	781	TACCTAAGATCTAGAGTCAAAATGTTTAAATTGCAAGGGCGTTTCAGATACATTTTATCCCA	840			
Dd	781	TACCTAAGATCTAGAGTCAAAATGTTTAAATTGCAAGGGCGTTTCAGATACATTTTATCCCA	840			
QY	841	TATCTTGATATACATTGGAGATAAAGCTGCTCAGAGTGTCACAAGGCATTTGGAGCAG	900			
Dd	841	TATCTTGATATACATTGGAGATAAAGCTGCTCAGAGTGTCACAAGGCATTTGGAGCAG	900			
QY	901	TTTGTGAAGCCGGAACAGCTTCATGGAGAAACCTCGTACAAAGTCAGCAAGTGTAAAAAG	960			
Dd	901	TTTGTGAAGCCGGAACAGCTTCATGGAGAAACCTCGTACAAAGTGTAAAAAG	960			
QY	961	ATGGLTTCAGCTTCAAGAGGTTCTATCATAGATCCTCTAATGTTCTTACACTTTCT	1020			
Dd	961	ATGGLTTCAGCTTCAAGAGGTTCTATCATAGATCCTCTAATGTTCTTACACTTTCT	1020			

```

QY 2101 GATGGTCCAGCTGCCAAGGCGAGCTGCCCTGCACCTCAGAAAAATCCCTTTGTAAGGCA 2160
Db 2101 GATGGTCCAGCTGCCAAGGCGAGCTGCCCTGCACCTCAGAAAAATCCCTTTGTAAGGCA 2160
QY 2161 AACGGTCTTCCTGGAAGTTGATGCTGCTCCTTTGCTGTCTCTCCAGAGACAAATC 2220
Db 2161 AACGGTCTTCCTGGAAGTTGATGCTGCTCCTTTGCTGTCTCTCCAGAGACAAATC 2220
QY 2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAAGCTCGACGGATGAATGAGTGCACCT 2280
Db 2221 TTAGAGACCTTCAGGCTTAGCAACAACTGAAAGCTCGACGGATGAATGAGTGCACCT 2280
QY 2281 GGAGCAGAGAGGGGCCCTCCGAGGAGCGGAGACCGGAGCTCAGCCTGGCAGCCCGGCC 2340
Db 2281 GGAGCAGAGAGGGGCCCTCCGAGGAGCGGAGACCGGAGCTCAGCCTGGCAGCCCGGCC 2340
QY 2341 GCCGAATCCCTGGAGGAGCCAGATGCGGCGCGCAAGCTTATCCCTTTAGTGGGGTTAA 2400
Db 2341 GCCGAATCCCTGGAGGAGCCAGATGCGGCGCGCAAGCTTATCCCTTTAGTGGGGTTAA 2400
QY 2401 TTTTAGCTTGCA 2412
Db 2401 TTTTAGCTTGCA 2412

RESULT 2
LOCUS AX786887 4659 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequence 22 from Patent WO0238744.
ACCESSION AX786887
VERSION AX786887.1 GI:32954188
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Lee,E.A., Hafalia,A.J., Yue,H., Lal,P.G., Yao,M.G., Lu,Y.,
Chawla,N.K., Warren,B.A., Lu,D.A., Baughn,M.R., Delegeane,A.M.,
Burford,N., Borowsky,M.L., Lee,S., Xu,Y., Griffin,J.A.,
Kallick,D.A., Gandhi,A.R., Arvizu,C.S., Ison,C.H., Tang,Y.T.,
Azinzai,Y., Elliott,V.S., Swarnakar,A., Ramkumar,J., Nguyen,D.B.,
Tribouley,C.M., Lo,T.P., Au-Young,J., Thangavelu,K. and Kearney,L.
Proteases
Patent: WO 0238744-A 22 16-MAY-2002;
Incyte Genomics, Inc. (US)
FEATURES
source
location/Qualifiers
1..4659
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/noe="Incyte ID No: 7997881CB1"

ORIGIN
Query Match 96.2%; Score 2321; DB 6; Length 4659;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGCTGTGTGGCGCGCGCGCGCGCGCGAGGGGATGGAGCGAGCGCGCGCGGT 60
Db 83 GGGCTGTGTGGCGCGCGCGCGCGCGCGAGGGGATGGAGCGAGCGCGCGCGGT 142
QY 61 CAGAGTTGAAACATGACCATGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 120
Db 143 CAGAGTTGAAACATGACCATGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTAT 202
QY 121 CAGATCAGCCTGGCAGCTCCGAGGAGCTTCACCTGGAGACATGGATGACGGTTCGCG 180
Db 203 CAGATCAGCCTGGCAGCTCCGAGGAGCTTCACCTGGAGACATGGATGACGGTTCGCG 262
QY 181 AGCTGGGTGCTGTCTTTCATGAATGATGTGCAAAATCACAACATTCATTTAGGACCA 240
Db 263 AGCTGGGTGCTGTCTTTCATGAATGATGTGCAAAATCACAACATTCATTTAGGACCA 322

```

```

QY 241 GTACCTGGTGTGTAGTTTATTCGAGTTCATCTGTACCTGTATAAATCAAAACCAATCACA 300
Db 323 GTACCTGGTGTGTAGTTTATTCGAGTTCATCTGTACCTGTATAAATCAAAACCAATCACA 382
QY 301 CAAAAGGATCAAGCCCTAGGTGATGGCATGCTCTCTCAGCAAGAAAGTTCTTTTCCATCT 360
Db 383 CAAAAGGATCAAGCCCTAGGTGATGGCATGCTCTCTCAGCAAGAAAGTTCTTTTCCATCT 442
QY 361 GAGAAGATTTTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 420
Db 443 GAGAAGATTTTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 502
QY 421 GGCAAATACCTGTTTGGCAATGCAGCACTGCGAGTGTAAACCTACACACCACTCTTTGCC 480
Db 503 GGCAAATACCTGTTTGGCAATGCAGCACTGCGAGTGTAAACCTACACACCACTCTTTGCC 562
QY 481 AATTACATGCTATCAGATGAACACTCCAAACATGTCATGAGAGGCTTTTGTATGATG 540
Db 563 AATTACATGCTATCAGATGAACACTCCAAACATGTCATGAGAGGCTTTTGTATGATG 622
QY 541 TGTACAATGCAAGCACATATTAACCGAGGCACTCAGTAATCCTGGGAGCTTTATTAACCA 600
Db 623 TGTACAATGCAAGCACATATTAACCGAGGCACTCAGTAATCCTGGGAGCTTTATTAACCA 682
QY 601 ATGTTTGTCTAATCAATGAGATGCGGCGGTATAGTAGGCACCTCCGTTTGGAAACCAAGAA 660
Db 683 ATGTTTGTCTAATCAATGAGATGCGGCGGTATAGTAGGCACCTCCGTTTGGAAACCAAGAA 742
QY 661 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGAGAGAGGCTTTGATGAGG 720
Db 743 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGAGAGAGGCTTTGATGAGG 802
QY 721 AGCAATAAATTAGACAGACACACCCAGGCGCACCTCTCTGTTGTGAGATTTGGAGGA 780
Db 803 AGCAATAAATTAGACAGACACACCCAGGCGCACCTCTCTGTTGTGAGATTTGGAGGA 862
QY 781 TACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTTCAGATACCTTTGATCCA 840
Db 863 TACCTAAGATCTAGAGTCAAAATGTTTAAATGCAAGGCGTTCAGATACCTTTGATCCA 922
QY 841 TATCTTGTATTAACATTTGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG 900
Db 923 TATCTTGTATTAACATTTGAGATAAAGGCTGCTCAGAGTGTCAACAGGCAATTTGGAGCAG 982
QY 901 TTTGTGAGCGCGGAAACAGCTTGCATGAGAGAAATCTCGTACAAGTGCAGCAAGTCTAAAAG 960
Db 983 TTTGTGAGCGCGGAAACAGCTTGCATGAGAGAAATCTCGTACAAGTGCAGCAAGTCTAAAAG 1042
QY 961 ATGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCT 1020
Db 1043 ATGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCT 1102
QY 1021 CTGAAACGTTTTCGCAAAATTTACCGGTGAAAAATGCTTAAGGATGTGAAATACCTGAG 1080
Db 1103 CTGAAACGTTTTCGCAAAATTTTACCGGTGAAAAATGCTTAAGGATGTGAAATACCTGAG 1162
QY 1081 TATCTTGTATTTGGGCCATATATGTTCTCAACCAACGAGAGCCAAATCTCTACAGTCTTG 1140
Db 1163 TATCTTGTATTTGGGCCATATATGTTCTCAACCAACGAGAGCCAAATGTTCTACAGTCTTG 1222
QY 1141 TATGAGTGTGGTCCACACTGGTTTTTAATGCCATGTGGCCATTACTTCTGTACATA 1200
Db 1223 TATGAGTGTGGTCCACACTGGTTTTTAATGCCATGTGGCCATTACTTCTGTACATA 1282
QY 1201 AAAGTACCAATGGCCTCTGGTATCAAAATGATGACTCCATGATCTACACAGTATTT 1260
Db 1283 AAAGTACCAATGGCCTCTGGTATCAAAATGATGACTCCATGATCTACACAGTATTT 1342
QY 1261 AGATCGGTACTCAGCCAAACAAAGCCCTATGTGCTCTTTTATATCAGGTCCCATGTGAAA 1320
Db 1343 AGATCGGTACTCAGCCAAACAAAGCCCTATGTGCTCTTTTATATCAGGTCCCATGTGAAA 1402

```

RESULT 3
AXI07851
LOCUS
DEFINITION
Sequence 2 from Patent WO0123589.
AXI07851
ACCESSION
VERSION
KEYWORDS
AXI07851.1 GI:13923242
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
Kapeller-Liberman, R.
2431, a novel human ubiquitin protease
PATENT: WO 0123589-A 2 05-APR-2001.
JOURNAL
MILLENNIUM PHARMACEUTICALS, INC. (US)
FEATURES
Location/Qualifiers
1..2347
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="23431 Ubiquitin protease"
61..>2147
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC37760.1"
/db_xref="GI:13923243"
/db_xref="REMTREMBL:CAC37760"
/translation="MTIVDKASRSSDPSEAYQNOPGSSEAVSPGMDAGSAWSGAVSSLNVSHTLSLGVPFGAVVSSSVSPKSKPKOQALGDGLAIPPOKVLPSEKICLKWOOTHRVGAGLQNLGNTCFANAALQCITTPPLANYMLSHESKTCHAEFCMCTMQAHTTQALSNPDGVTKFMFVNEMRIARHFPGNQEDAEFLQTVDMAOKACLINGSKLDRHTTPTLVCI1FGGYLRSCVKLNCKGVSDTFDYLDITLEIKAAQSVNKALQSFPVEQLDNGSGYKCSKKWMPASKRFTHRSNNVLTLKRPFANTGGIKIAKOVKYPEYLDRPYMSQPNSKEPIVVVLAVLVHTGNCAGHYCYIKASNLGTQMNDISIVSTSIIRSVLSQAQVILPWRSHDKNGELTHPTSPQSSRPPIVISORVVTNOAAPGFQGPLDSHMKNPHHLTGTPCLKDTPESSMSPPNGSNSVRASPASVONQSWNVNRSSVIPHPHKQKITISHNKLIPVRCSSQPNLHSNLENFTXPVPSITTNSAVOSTSNASTYSVSKVTKPIPRSESCQPVMNGSKLNSVLYPYAESDESDDEESKGLCGENIGTTVSSHSPQDAEAPTHELQEPMTLNGANSADSDPKENGSLAPDGASCQGOPALHSENPAFKANGLPGLKMPALLSLPEDKILETFRLSNKLKGSTDMSAPGAERGPPEDRDAEPQGPSRAESLEE"
ORIGIN
Query Match 94.8%; Score 2287; DB 6; Length 2347;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2337; Conservatives 0; Mismatches 1; Indels 0; Gaps 0;
QY 22 CGCGCGCCGAGGGGATGGAGCAGCGCGAGCGCGGTCTAGAGTTGAAACAATGACCATA 81
DB 10 CGCGCGCCGAGGGGATGGAGCAGCGCGGTCTAGAGTTGAAACAATGACCATA 69
QY 82 GTTCACAAAAGCTTCTGAATCTTCAGACCCCATCAGCCTATCAGAATCAGCCTGGCAGCTCC 141
DB 70 GTTGACAAAGCTTCTGAATCTTCAGACCCCATCAGCCTATCAGAATCAGCCTGGCAGCTCC 129
QY 142 GAGGCAGTCTCACTGGAGACATGATGCCAGTTCGCCAGCTGGGGTGCTGTGTCTTCA 201
DB 130 GAGGCAGTCTCACTGGAGACATGATGCCAGTTCGCCAGCTGGGGTGCTGTGTCTTCA 189
QY 202 TTGAATGATGTCAAATCAACACACTTCCTTTAGGACAGGTACCTGGTGCTGTAGTTTAT 261
DB 190 TTGAATGATGTCAAATCAACACTTCCTTTAGGACAGGTACCTGGTGCTGTAGTTTAT 249
QY 262 TCGAGTTCACTGTACTCTGATAAATCAAACCATCAACAAAGGATCAAGCCCTAGGT 321
DB 250 TCGAGTTCACTGTACTCTGATAAATCAAACCATCAACAAAGGATCAAGCCCTAGGT 309
QY 322 GATGGCATCGTCTCTCCACAGAAAGTCTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGG 381
DB 310 GATGGCATCGTCTCTCCACAGAAAGTCTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGG 369
QY 382 CAACAAACATCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTTGCCAAT 441

Db	370	CAACAACTCATAGAGTTGGAGCTGGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCCAAT	429
Qy	442	GCAGCACTGCAGTGTTTAAACCTACACACCACTCTTGCCAAATACATGCTATCACATGAA	501
Db	430	GCAGCACTGCAGTGTTTAAACCTACACACCACTCTTGCCAAATACATGCTATCACATGAA	489
Qy	502	CACCTCCAAAACATGTCATGCAGAGGCTTTTGTATGATGTGTA CAATGCAGACATATT	561
Db	490	CACCTCCAAAACATGTCATGCAGAGGCTTTTGTATGATGTGTA CAATGCAGACATATT	549
Qy	562	ACCAGGCACTCAGTAATCCTGGGGAGCTTATTAAACCAATGTTTGTATCAATGAGATG	621
Db	550	ACCAGGCACTCAGTAATCCTGGGGAGCTTATTAAACCAATGTTTGTATCAATGAGATG	609
Qy	622	CGGCGTATAGTAGGCCACTCCGTTTGTGGAAACCAAGAAGATGCCATGAATTCCTTCAA	681
Db	610	CGGCGTATAGTAGGCCACTCCGTTTGTGGAAACCAAGAAGATGCCATGAATTCCTTCAA	669
Qy	682	TACACTGTTGATGTCATGCGAAAGCATGCTTGAATGGCGCAATAAAATAGACAGACAC	741
Db	670	TACACTGTTGATGTCATGCGAAAGCATGCTTGAATGGCGCAATAAAATAGACAGACAC	729
Qy	742	ACCAGGCCCCACTCTGTTGTCAGATATTTGGAGGATACCTTAAGATCTAGAGTCAAA	801
Db	730	ACCAGGCCCCACTCTGTTGTCAGATATTTGGAGGATACCTTAAGATCTAGAGTCAAA	789
Qy	802	TGTTTAAATTCGAAGGGCGTTTCAGATACATTTTGTATCCATATCTTGATATAACAATTCGAG	861
Db	790	TGTTTAAATTCGAAGGGCGTTTCAGATACATTTTGTATCCATATCTTGATATAACAATTCGAG	849
Qy	862	ATAAAGCGTGTCTAGAGTGTCACAAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTT	921
Db	850	ATAAAGCGTGTCTAGAGTGTCACAAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTT	909
Qy	922	GATGGAGAAACTCGTACAAAGTCGAGCAAGTGTA AAAAGATGGTTCAAGCTTCAAAGAGG	981
Db	910	GATGGAGAAACTCGTACAAAGTCGAGCAAGTGTA AAAAGATGGTTCAAGCTTCAAAGAGG	969
Qy	982	TTCACTATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTGTGAAATTTT	1041
Db	970	TTCACTATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTGTGAAATTTT	1029
Qy	1042	ACGGTGGAAAAATTGCTAAAGATGTGAATAACCTGAGTATCTTGATATTCGGCCATAT	1101
Db	1030	ACGGTGGAAAAATTGCTAAAGATGTGAATAACCTGAGTATCTTGATATTCGGCCATAT	1089
Qy	1102	ATGTCCTCAAACCAACGAGAGCCAAATGTCTACGTCTTGATGTCAGTGCTGGTCCACACT	1161
Db	1090	ATGTCCTCAAACCAACGAGAGCCAAATGTCTACGTCTTGATGTCAGTGCTGGTCCACACT	1149
Qy	1162	GGTTTAAATTTGCCATGCTGGCCATTACTCTGTCTACATAAAGCTAGCAATGCCCTCTGG	1221
Db	1150	GGTTTAAATTTGCCATGCTGGCCATTACTCTGTCTACATAAAGCTAGCAATGCCCTCTGG	1209
Qy	1222	TATCAAATGAATGACTCCAATTTGATCTACACAGTGATATTAGATCGGTACTCAGCCCAACA	1281
Db	1210	TATCAAATGAATGACTCCAATTTGATCTACACAGTGATATTAGATCGGTACTCAGCCCAACA	1269
Qy	1282	GCCTATGTGCTCTTTTATATCAGGTCCTCCATGATGTGAAAAATGGAGGTGAACCTTACTCAT	1341
Db	1270	GCCTATGTGCTCTTTTATATCAGGTCCTCCATGATGTGAAAAATGGAGGTGAACCTTACTCAT	1329
Qy	1342	CCACCCATAGCCCCGCCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTCAACC	1401
Db	1330	CCACCCATAGCCCCGCCAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTCAACC	1389
Qy	1402	AACAAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTCTCATCATGATAAAGAAT	1461
Db	1390	AACAAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTCTCATCATGATAAAGAAT	1449
Qy	1462	CCACCTCATTTAAATGGGACTGGACCAATTGAAAGACACGCCAAGAGTTCCTATGTCGAGT	1521

Db	1450	CCACCTCACTTAAATGGGACTGGACCAATTGAAAGACACGCGAAGCACTTCCATGTCGAGT	1509
Qy	1522	CCTAACGGGAATTCAGGTGTCAACAGGGCTAGTCTGCTTAATGCTTCAGCTTCTGTGTCAA	1581
Db	1510	CCTAACGGGAATTCAGGTGTCAACAGGGCTAGTCTGCTTAATGCTTCAGCTTCTGTGTCAA	1569
Qy	1582	AATGTGTCAGTTAATAGGTCTCTAGTGATCCGAGAACATCCTAAGAAACAAAAAATTACA	1641
Db	1570	AATGTGTCAGTTAATAGGTCTCTAGTGATCCGAGAACATCCTAAGAAACAAAAAATTACA	1629
Qy	1642	ATCAGTATTACACACAGATTGCCCTGTTTCGCCAGTGTGTCAGTCTCAACCTAACCTTCATAGT	1701
Db	1630	ATCAGTATTACACACAGATTGCCCTGTTTCGCCAGTGTGTCAGTCTCAACCTAACCTTCATAGT	1689
Qy	1702	AATTCTTTGGAGAACCTTACCAGGCCGTTCCCTCTTCTACCATTAACCAATCTGCGAGTA	1761
Db	1690	AATTCTTTGGAGAACCTTACCAGGCCGTTCCCTCTTCTACCATTAACCAATCTGCGAGTA	1749
Qy	1762	CAGTCTACCTCGAACGCAATCTACGATGTCAAGTCTAGTTTCTAGTAAAGTAAACAAACCGATCCCC	1821
Db	1750	CAGTCTACCTCGAACGCAATCTACGATGTCAAGTCTAGTTTCTAGTAAAGTAAACAAACCGATCCCC	1809
Qy	1822	CGCAGTGAATCTGCTCTCCAGCCGCTGATGAATGCGAAATCBAAGCTGAAGTCTCAGCGTG	1881
Db	1810	CGCAGTGAATCTGCTCTCCAGCCGCTGATGAATGCGAAATCBAAGCTGAAGTCTCAGCGTG	1869
Qy	1882	CTGGTGCCCTATGGCGCGAGTCTCTCTGAGGATCTCTGACGAGAGTCTCAAGGGGCTGGGC	1941
Db	1870	CTGGTGCCCTATGGCGCGAGTCTCTCTGAGGATCTCTGACGAGTCTCAAGGGGCTGGGC	1929
Qy	1942	AAGGAGAAATGGGATTTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCGAAGAT	2001
Db	1930	AAGGAGAAATGGGATTTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCGAAGAT	1989
Qy	2002	GAGGAGGCCACTCGCAGCAGCTTCAAGAACCAATGACCTTAAACGCTGCTAATAGTGCA	2061
Db	1990	GAGGAGGCCACTCGCAGCAGCTTCAAGAACCAATGACCTTAAACGCTGCTAATAGTGCA	2049
Qy	2062	GACAGCGACAGTGACCCGAAAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGC	2121
Db	2050	GACAGCGACAGTGACCCGAAAGAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGC	2109
Qy	2122	CAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTCGTGAAGTTG	2181
Db	2110	CAGCTGCCCTGCACCTCAGAAAAATCCCTTTGCTTAAGGCAAAACGGTCTTCTCGTGAAGTTG	2169
Qy	2182	ATGCTGCTCTTTGCTGTCTCTCCAGAAAGCAAAATCTTAGAGACCTTTCAGGCTTAGC	2241
Db	2170	ATGCTGCTCTTTGCTGTCTCTCCAGAAAGCAAAATCTTAGAGACCTTTCAGGCTTAGC	2229
Qy	2242	AACAACTGAAAGGCTCGACGGATGAATCAGTGACCTGGAGCGAGAGGGGCCCTCCC	2301
Db	2230	AACAACTGAAAGGCTCGACGGATGAATCAGTGACCTGGAGCGAGAGGGGCCCTCCC	2289
Qy	2302	GAGGACCGGACCGCGAGCTCAGCCTGGCAGCCCGCGCGCGAAATCCCTTGGAGGAGC	2359
Db	2290	GAGGACCGGACCGCGAGCTCAGCCTGGCAGCCCGCGCGCGAAATCCCTTGGAGGAGC	2347
RESULT 4			
AX882263			
LOCUS	AX882263	3660 bp	DNA
DEFINITION	Sequence 17168 from Patent EP1074617.		
ACCESSION	AX882263		
VERSION	AX882263.1	GI:40037087	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	AX882263	linear	PAT 17-DEC-2003
	Sequence 17168 from Patent EP1074617.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1. Ochi, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J., Ichi, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.		

TITLE Primers for synthesising full-length cDNA and their use

JOURNAL Patent: EP 1074617-A 17168 07-FEB-2001;
Research Association for Biotechnology (JP)

FEATURES
Location/Qualifiers
source

1..3660

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

CDS

68..53658

/note="unnamed protein product"

/codon_start=1

/protein_id="CAE92364.1"

/db_xref="GI:40037088"

/translation="MTIVKASSSDPSAYQNGPSSEAYSPGMDAGSASGAVSL
NDVSNHTLSLGPVGVAVVSSSVDPKSKSPQDQALGDGIAPQKQLPFPSEKICLK
WQTHRVLGAGLNGNTFANAALQCLTYTPLANYMLSHEHSKTCBAEGFCWMCITQ
AHITQALSNPDVTKPMFVNMWRIRARHEFQNEADAEFLQYTVDMQAKCLNSG
KLDRHTQATLVQCI FGGLRSRVKLNCKGVSDTFDPYLDITLEIKAAQSVNKALQ
FVKEQLDGENSKCSKCKWVPASKRFTIHRSSNVLTLSLKRFANFTGGKIAKVQY
PEYLDIRPYMSQNGEPIVYLVAVLHTGFCHAGHYFCYIKASNGLWYQMDSDIVS
TSDIRSVLSQQAIVLYFIRSHDVKNKGELTHPTSPGSPRPVVISORVTVNKQAAPG
FIQGPQSPHMKMPPHLLNGTGLPKDTSSSPGNSPNSVNRASPVNASVQNWQSVN
RSSVPIPEHPKQKQITISHNKLVRQCQSPNLHNSLENTKPVSSITINSVQST
SNASTWSVSKVTKPI PRSSCSOPVMNGSKLNSVLVPYGAESSDEESKGLGK
ENGLTIVSSHSPQDAEDBEATPEHLEQPTTLKNGANSADSDPKENGLAPGASQC
EQPALHSENFPAXANGLPGLMPAPILSLPEDKILETFRLSNKLKSTDEMSAPGER
GPEDRDAGEQSPAESLEEDPAAGLSLSTKAKPPRDPDPTATKEGWEAMAVAP
EEPPPSADGIVGTAPDPLCDGLSDASPLSQDAKMIAGSPRDSALAEAPEGLS
PAPPSREEQEDPLLHPSPGSDHARDQDPSQSLGAPAEARPPAPVLDWAPAGHPG
DAPSPERVEADAAPGAPSPAKESIGSLRKVDGRHYRSTRERSSGSPARESRK
TEGHRHRRRTCTPERDQRHAPPEHPHGDRLSPGERRSLGRCSHHHSHRHSRVBL
DWYRHVYTEGEMGRKREKFPDRPWDRCRYHDRVLYAARDWKPHGPHGERHACGL
HERPHKDNHRRRCCEPARERHRPSSPRAGAPHALAPHPDRPSHRTALVAGDNCN
LSDRFHEHNGKSKRRHSDSVENSDSHVEKKARRSEKQDPLEPFPKAKHKK"

ORIGIN

Query Match 93.9%; Score 2265; DB 6; Length 3660;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2365; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	6	GTGTGCGCGCGCGCGCGCGCGGAGGAGGATGGAGCGAGCGCCGAGCGGCTCAGAG	65
Db	1	GTGTGCGCGCGCGCGCGCGCGGAGGAGGATGGAGCGAGCGCGCGGCTCAGAG	60
Qy	66	TTGACAAATGACCAATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCTATCAGAA	125
Db	61	TTGACAAATGACCAATAGTTGACAAAGCTTCTGAATCTTCAGACCCATCAGCTATCAGAA	120
Qy	126	TCAGCTGGCAGCTCCGAGGCGAGTCTCACCTGGGAGACATGATGCGAGGTTCTGCCAGCTG	185
Db	121	TCAGCTGGCAGCTCCGAGGCGAGTCTCACCTGGGAGACATGATGCGAGGTTCTGCCAGCTG	180
Qy	186	GGGTGCTGTGTCTTCAATGAATGATGTGTCAAATACACACATTTCTTTTAGGACCGAGTACC	245
Db	181	GGGTGCTGTGTCTTCAATGAATGATGTGTCAAATACACACATTTCTTTTAGGACCGAGTACC	240
Qy	246	TGGTGTGTGTATTTATTCGAGTTTCATCTGTACCTGATATAATCAAAACCATCACCACAAAA	305
Db	241	TGGTGTGTGTATTTATTCGAGTTTCATCTGTACCTGATATAATCAAAACCATCACCACAAAA	300
Qy	306	GGATCAAGCCCTAGTGATGCGATCGCTCCCTCCACAGAAAAGTTCTTTTCCCATCTGAGAA	365
Db	301	GGATCAAGCCCTAGTGATGCGATCGCTCCCTCCACAGAAAAGTTCTTTTCCCATCTGAGAA	360
Qy	366	GATTGTCTTAAGTGGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAA	425
Db	361	GATTGTCTTAAGTGGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAA	420
Qy	426	TACCTGTTTTCGAATGAGCACTGTCAGTGTTTAACTTACACACACACCTCTTGGCAATTA	485
Db	421	TACCTGTTTTCGAATGAGCACTGTCAGTGTTTAACTTAACTTACACACACACCTCTTGGCAATTA	480
Qy	486	CATGCTATCACATGAACACTCCAAAACATGTGATGCAAGAGGCTTTTGTGATGATGTATC	545

Db	481	CATGCTATCACATGAACACTCCAAAAACATGCTCATGCAAGAGGCTTTTGTATGATGTATC	540
Qy	546	AATGCAAGCACAATATACCCAGGCACTCAGTAATCTCTGGGAGCGTTATTTAAACCAATGTT	605
Db	541	AATGCAAGCACAATATACCCAGGCACTCAGTAATCTCTGGGAGCGTTATTTAAACCAATGTT	600
Qy	606	TGTCATCAATGAGATGCGGGGTATAGCTAGGCACTCCGCTTTTGGAAACCAAGAAGATGC	665
Db	601	TGTCATCAATGAGATGCGGGGTATAGCTAGGCACTCCGCTTTTGGAAACCAAGAAGATGC	660
Qy	666	CCATGAATTCCTTCAATACACTGTGTGATGCTATGCAAGAGCACTGTTGAATGSCAGCAA	725
Db	661	CCATGAATTCCTTCAATACACTGTGTGATGCTATGCAAGAGCACTGTTGAATGSCAGCAA	720
Qy	726	TAAATTTAGACAGACACACCCAGGCCACCTCTTGTGTTTGCAGATATTTGGAGGATACCT	785
Db	721	TAAATTTAGACAGACACACCCAGGCCACCTCTTGTGTTTGCAGATATTTGGAGGATACCT	780
Qy	786	AAGATCTAGAGTCAATATGTTTAAATTTCAAGGGCTTTTCAGATACCTTTTGATCCATATCT	845
Db	781	AAGATCTAGAGTCAATATGTTTAAATTTCAAGGGCTTTTCAGATACCTTTTGATCCATATCT	840
Qy	846	TGATATAACATTTGGAATAAGGCTGCTCAGAGTGTCAAAGGCATTTGGAGCAGTTTGT	905
Db	841	TGATATAACATTTGGAATAAGGCTGCTCAGAGTGTCAAAGGCATTTGGAGCAGTTTGT	900
Qy	906	GAAGCCGGACACGCTTGATGAGAGAACTCGTCAAGTCGAGCAAGTGTAAAAAGATGGT	965
Db	901	GAAGCCGGACACGCTTGATGAGAGAACTCGTCAAGTCGAGCAAGTGTAAAAAGATGGT	960
Qy	966	TCCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCTCTGAA	1025
Db	961	TCCAGCTTCAAGAGGTTTCACTATCCATAGATCCTCTAAATGTTCTTACACTTTCTCTGAA	1020
Qy	1026	ACGTTTTGCAAAATTTTACCAGTGGAAAAATTTGCTAAGGATGTGAAATACCCCTGAGTATCT	1085
Db	1021	ACGTTTTGCAAAATTTTACCAGTGGAAAAATTTGCTAAGGATGTGAAATACCCCTGAGTATCT	1080
Qy	1086	TGATATTCGGCCATATATGCTCAACCAAGGAGAGCCAAATGCTACGCTCTGTGATGC	1145
Db	1081	TGATATTCGGCCATATATGCTCAACCAAGGAGAGCCAAATGCTACGCTCTGTGATGC	1140
Qy	1146	AGTCTCGTCCACACTGCTTTTAAATGGCATGCTGGCCATTTACTTCTGTACATAAAAGC	1205
Db	1141	AGTCTCGTCCACACTGCTTTTAAATGGCATGCTGGCCATTTACTTCTGTACATAAAAGC	1200
Qy	1206	TAGCAATGGGCTCTGTATCAAATGAATGACTCCATTTGATCTACAGTGTATTAGATC	1265
Db	1201	TAGCAATGGGCTCTGTATCAAATGAATGACTCCATTTGATCTACAGTGTATTAGATC	1260
Qy	1266	GGTACTCAGCCCAACAGCCCTATGCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGG	1325
Db	1261	GGTACTCAGCCCAACAGCCCTATGCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGG	1320
Qy	1326	AGGTGAATTTACTCATCCCAATAGCCCGGCGAGTCTCTCCCGCCCGCCGTCATCAG	1385
Db	1321	AGGTGAATTTACTCATCCCAATAGCCCGGCGAGTCTCTCCCGCCCGCCGTCATCAG	1380
Qy	1386	TCAGCGGTTGTTCACCAACAAAGGCTGCGCCAGGCTTTTATCGGACACAGCTTCCCTC	1445
Db	1381	TCAGCGGTTGTTCACCAACAAAGGCTGCGCCAGGCTTTTATCGGACACAGCTTCCCTC	1440
Qy	1446	TCACATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCATTTGAAAGACACGCCAAG	1505
Db	1441	TCACATGATAAAGAAATCCACCTCACTTAAATGGGACTGGACCATTTGAAAGACACGCCAAG	1500
Qy	1506	CAGTTCCATGTGAGTCTCTAACCGGAAATTCAGTGTCCAGAGGCTAGTCTCTGTTAATGC	1565
Db	1501	CAGTTCCATGTGAGTCTCTAACCGGAAATTCAGTGTCCAGAGGCTAGTCTCTGTTAATGC	1560
Qy	1566	TTGAGCTTCTGTCCAAACTGGTTCAGTTAATAGTCTCTCAGTGATCCAGACACATCTTAA	1625

Qy	726	TAATTTAGACAGACACACCGAGCCACCACTCTTGTTTGTCCAGATATTGGAGGTACCT	785
Db	721	TAATTTAGACAGACACACCGAGCCACCACTCTTGTTTGTCCAGATATTGGAGGTACCT	780
Qy	786	AAGATCTPAGAGTCAAAATGTTTAAATTTGCAAGGCGTTTCAGATACCTTTTGATCCATATCT	845
Db	781	AAGATCTPAGAGTCAAAATGTTTAAATTTGCAAGGCGTTTCAGATACCTTTTGATCCATATCT	840
Qy	846	TGATATACCAATTCGAGATAAAGCTCCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGT	905
Db	841	TGATATACCAATTCGAGATAAAGCTCCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGT	900
Qy	906	GAAGCCGGAACAGCTTGATCGAGAAAACCTCGTACAAGTCGACCAAGTGTAAAAAGATGCT	965
Db	901	GAAGCCGGAACAGCTTGATCGAGAAAACCTCGTACAAGTCGACCAAGTGTAAAAAGATGCT	960
Qy	966	TCCAGCTTCAAAAGAGTTCACTATCCATAGATCCTCTAATGTCTTACATTTTACATTTCTCTGAA	1025
Db	961	TCCAGCTTCAAAAGAGTTCACTATCCATAGATCCTCTAATGTCTTACATTTTCTCTGAA	1020
Qy	1026	ACGTTTTGCAAAATTTTACCGGTGCAAAAATGCTAAAGGATGCAAAATACCTCGAGTATCT	1085
Db	1021	ACGTTTTGCAAAATTTTACCGGTGCAAAAATGCTAAAGGATGCAAAATACCTCGAGTATCT	1080
Qy	1086	TGATATTCGGCCATATATGTCTCAACCCACGGAGAGCCAAATGTCTACGTCTTGATGC	1145
Db	1081	TGATATTCGGCCATATATGTCTCAACCCACGGAGAGCCAAATGTCTACGTCTTGATGC	1140
Qy	1146	AGTGTGTGCCACTGCTGGTTTTTAATTCGCAATGCTGGCCATTACTTCTGCTACATAAAGC	1205
Db	1141	AGTGTGTGCCACTGCTGGTTTTTAATTCGCAATGCTGGCCATTACTTCTGCTACATAAAGC	1200
Qy	1206	TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTGTATCTACCAGTGATATAGATC	1265
Db	1201	TAGCAATGGCCCTCTGGTATCAAAATGAATGACTCCATTGTATCTACCAGTGATATAGATC	1260
Qy	1266	GGTACTCAGCCAAACAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGAAAAATGG	1325
Db	1261	GGTACTCAGCCAAACAGCCTATGTGCTCTTTTATATCAGGTCCCATGATGAAAAATGG	1320
Qy	1326	AGGTGAACCTTACTCATCCACCCATAGCCCGGCGCAGTCTCTTCCCGCCCCGTCATCAG	1385
Db	1321	AGGTGAACCTTACTCATCCACCCATAGCCCGGCGCAGTCTCTTCCCGCCCCGTCATCAG	1380
Qy	1386	TCAGCGGTGTGTCAACCAACAAAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTC	1445
Db	1381	TCAGCGGTGTGTCAACCAACAAAGGCTGGCCAGGCTTTATCGGACCAAGCTTCCCTC	1440
Qy	1446	TCACATGATAAAGAAATCCACCTCACTTAAATPGGAATGGACCAATGAAAGACACGCCAAG	1505
Db	1441	TCACATGATAAAGAAATCCACCTCACTTAAATPGGAATGGACCAATGAAAGACACGCCAAG	1500
Qy	1506	CAGTTCCATGTCGAGTCCCTAACGGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGC	1565
Db	1501	CAGTTCCATGTCGAGTCCCTAACGGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGC	1560
Qy	1566	TTACGCTTCTGTCCAAAACCTGGTCAGTTTAAATAGTGTCTCAGTGATCCAGAAACATCCTAA	1625
Db	1561	TTACGCTTCTGTCCAAAACCTGGTCAGTTTAAATAGTGTCTCAGTGATCCAGAAACATCCTAA	1620
Qy	1626	GAAACAAAAAATTAACAATCAGTATTTCAACAAGATGCTGTTCCGCAAGTGTCAAGTCTCA	1685
Db	1621	GAAACAAAAAATTAACAATCAGTATTTCAACAAGATGCTGTTCCGCAAGTGTCAAGTCTCA	1680
Qy	1686	ACCTAACCTTCATAGTAAATCTTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT	1745
Db	1681	ACCTAACCTTCATAGTAAATCTTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCAT	1740
Qy	1746	TACCAATTCGAGGTACAGTCTTACCTCGAAACGCACTTACGATGTCAAGTGTCTAGTAAAGT	1805
Db	1741	TACCAATTCGAGGTACAGTCTTACCTCGAAACGCACTTACGATGTCAAGTGTCTAGTAAAGT	1800
Qy	1806	ABACAAACCGATTCGCCCGCAGTGAAATCTTGCTCCAGCCCGTGTGATGAATGCAAAATCCAA	1865

Db	1801	AA	CAAAACCCAGTCCC	CGCAGTGAATCTCTGCTCC	CAGCCGCGTGAATGGCAAAATCCAA	1861
Qy	1866	GC	TGAATCTCAGCGT	GCTGGTGCCTTATGGCGCCGAGTCTCT	TGAGGACTCTTGACGAGGA	1925
Db	1861	GCT	TGAATCTCAGCGT	GCTGGTGCCTTATGGCGCCGAGTCTCT	TGAGGACTCTTGACGAGGA	1920
Qy	1926	GT	CAAAAGGGCTGGG	CAAGGAGAAATGGATTTGGTACGATTTG	TGAGTCCCACTCTCCCG	1985
Db	1921	GT	CAAAAGGGCTGGG	CAAGGAGAAATGGATTTGGTACGATTTG	TGAGTCCCACTCTCCCG	1980
Qy	1986	CCA	GATGCCGAAGTCA	GAGGAGCCACTCCGACAGAGCTTCAAGAA	ACCCATGACCCCTAAA	2045
Db	1981	CCA	GATGCCGAAGTCA	GAGGAGCCACTCCGACAGAGCTTCAAGAA	ACCCATGACCCCTAAA	2040
Qy	2046	CGT	GTCTAATAGTGC	ACAGACGACAGTACCCGAAAGAAACGGCT	TAGCGCTGATGG	2105
Db	2041	CGT	GTCTAATAGTGC	ACAGACGACAGTACCCGAAAGAAACGGCT	TAGCGCTGATGG	2100
Qy	2106	TGC	CAGTGCCAAGCC	CAGCGCTGCCCTGCACTCAGAAATCCCTTT	GCTTAAGCGCAACGG	2165
Db	2101	TGC	CAGTGCCAAGCC	CAGCGCTGCCCTGCACTCAGAAATCCCTTT	GCTTAAGCGCAACGG	2160
Qy	2166	TC	TTCTTGGAAATG	TGATGCTGCTCTTGGTGTCTCTCCCAAGAC	AAAAATCTTAGA	2225
Db	2161	TC	TTCTTGGAAATG	TGATGCTGCTCTTGGTGTCTCTCCCAAGAC	AAAAATCTTAGA	2220
Qy	2226	GAC	CTTCAGGCTTAG	CAACAACTGAAAGGCTCGACGGATGAAATG	AGTGCACCTTGAGC	2285
Db	2221	GAC	CTTCAGGCTTAG	CAACAACTGAAAGGCTCGACGGATGAAATG	AGTGCACCTTGAGC	2280
Qy	2286	AG	AGAGGGGCGCTCT	CGAGGACCGGAGCGGAGCTCAGCTTGGC	AGCCCGCCGCGGA	2345
Db	2281	AG	AGAGGGGCGCTCT	CGAGGACCGGAGCGGAGCTCAGCTTGGC	AGCCCGCCGCGGA	2340
Qy	2346	AT	CCCTGGAGGAGC	GCAGATGGGCCGC	2372	
Db	2341	AT	CCCTGGAGGAGC	GCAGATGGGCCGC	2367	
RESULT 6						
AK022759						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						

etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES

source

Location/Qualifiers

1..3660

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="NT2RP1000522"

/cell_line="NT2"

/cell_type="teratocarcinoma"

/clone_lib="NT2RP1"

/note="cloning vector: pUC19FL3-mRNA from NT2 neuronal precursor cells after 48-hours retinoic acid (RA) induction."

68..>3660

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB14232.1"

/db_xref="GI:10434352"

/translation="MTIVDKASESDPSAVONOPGSSSEAVSPGMDAGASWCAVSSL
NDVSNHTLSLGPVPGAVVSSSSVPDKSPKQKQALGDGIAPPQKVLFPSEKILK
WQOTHEVAGLONLNTCFANAALQCLTTPPLANTMLSEHSKTCCHAGFCMMCTMQ
AHTQALSNPQVIMKMFVINEMRIRARFQNGEDAHEFLQYTDVDMQKCLNGSN
KIDRHTQATLVCQIFGGYLSRVKCLNGKVSDDTDPYLDITLEIKAAQSVNKALEQ
FVKPEQLDGENSYKCSKCKMVPASXRFTHRSSNVLTLILKRFANFTGKIAKDVY
PYLDIRPYMSQPNQGEPIVVVLYAVLVHTGCHAGHYCYIKASNLWYQWDSIVS
TSDIRSLQQAIVLFIIRSHDVKNGELTHPTHPGQSPPRVISQRVVTNQAAAPG
FIGPOLPMHIMKNPHLNGTGPLKDTPPSSMSPNNGSSVNRPAISVASVONMSVN
RSVDEPHKPKQITILSHNKLPVROCOQPNLHNSLENPTKPVPSSTITNSAVOST
SNASTVSSKVTIPIRSESCSQPMYNGSKLNSVLVYGAESSEDSDEESGLGK
ENGIETIVSHSPGDAEAEATPHELOPTTLINGANSADSDPKENGLAPDGASQ
GQALHSENFPKANGLPKLPAPLLSLPEDIKLFTRLSNKLKSTDEMSAPGNER
GPEPDADBPQSPAAESLEPDAAGLSSTKAPPPDPDPTPATKEGAWEMAVAP
BEPPSAGEDI VGDTPAPDLCPGSLTGDASLSLQDAKMIAGEPDPSALAPAEGLS
PAPPARSEPCQPIVLLPSPGHDHADAQPSQSLGAPEAAERPPAPVLDMAPAHPEG
DAEPSGERVEDAAPKAPFPQSPAKEKISLRKVRDGHYRSRRSSSGEPARESRK
TEGHRHRTCTPRERDRHAPHHGPHGDRLLSPGERRSLSGRSHHSRHSRSGVEL
DWVRHHYTEGREGREKFPYDPRWRDRCYVDRYALVAAEDMKFPFGHGRHERAGL
HERPKDNRGRGCEPAREPERPSPRAGAPHALAPHPDRFSDHRTALVAGDNCN
LSDRHEHENGKSRVRHDSVENSISHVEKKARRSEQKDLPEPKAKKHKS"

ORIGIN

Query Match

Best Local Similarity 99.9%; Score 2265; DB 9; Length 3660;

Matches 2365; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	426	TACCTGTTTTGCCAATGACGACCTGAGTGTAACTTACACACACCTCTTGGCAATTA	485
DB	421	TACCTGTTTTGCCAATGACGACCTGAGTGTAACTTACACACACCTCTTGGCAATTA	480
QY	486	CATGCTATCATGAGAACACCTCCAAAACATGTCGAGAAAGGCTTTTGTATGATGTGTAC	545
DB	481	CATGCTATCATGAGAACACCTCCAAAACATGTCGAGAAAGGCTTTTGTATGATGTGTAC	540
QY	546	AATGCAAGACATATTACCCAGGACCTCAGTAATCTCTGGGGAGCGTTATTAACCAATGTT	605
DB	541	AATGCAAGACATATTACCCAGGACCTCAGTAATCTCTGGGGAGCGTTATTAACCAATGTT	600
QY	606	TGTCATCAATGAGATCGGGGTATAGCTAGGCACCTCCGTTTTGGAAACCAAGAGATGC	665
DB	601	TGTCATCAATGAGATCGGGGTATAGCTAGGCACCTCCGTTTTGGAAACCAAGAGATGC	660
QY	666	CCATGAATTCCTTCAATACACTGTTGATGCTATGCGAAAGCATGCTTGAATGGCAGCAA	725
DB	661	CCATGAATTCCTTCAATACACTGTTGATGCTATGCGAAAGCATGCTTGAATGGCAGCAA	720
QY	726	TAAATTTAGACAGACACACCCAGGCCACCACTCTTGTGTTGTCAGATATTTGGAGATACCT	785
DB	721	TAAATTTAGACAGACACACCCAGGCCACCACTCTTGTGTTGTCAGATATTTGGAGATACCT	780
QY	786	AGATCTAGAGTCAAAATGTTTAAATTCGAAGGGCGTTTCAGATACCTTTTGATCCATATCT	845
DB	781	AGATCTAGAGTCAAAATGTTTAAATTCGAAGGGCGTTTCAGATACCTTTTGATCCATATCT	840
QY	846	TGATATAAATTTGGAGATAAAGCTCTCAGAGTGTCAACAAGGCATTTGGAGCAGTTGT	905
DB	841	TGATATAAATTTGGAGATAAAGCTCTCAGAGTGTCAACAAGGCATTTGGAGCAGTTGT	900
QY	906	GAAGCGGAACAGCTTGATGGAGAAATCTGTACAGTGCAGCACTGTAAAGATGCT	965
DB	901	GAAGCGGAACAGCTTGATGGAGAAATCTGTACAGTGCAGCACTGTAAAGATGCT	960
QY	966	TCCAGGTTCAAAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCTCGAA	1025
DB	961	TCCAGGTTCAAAAGAGGTTCACTATCCATAGATCTCTTAATGTTCTTACACTTTCTCGAA	1020
QY	1026	ACGTTTTCGAAATTTACCGGTGGAAAAATTCGTAAGGATGTGAANAATCCCTGAGTATCT	1085
DB	1021	ACGTTTTCGAAATTTACCGGTGGAAAAATTCGTAAGGATGTGAANAATCCCTGAGTATCT	1080
QY	1086	TGATATTCGGCCCATATATGTCCTCAACCCAGGAGCCAAATGCTAGCTCTTGTATGC	1145
DB	1081	TGATATTCGGCCCATATATGTCCTCAACCCAGGAGCCAAATGCTAGCTCTTGTATGC	1140
QY	1146	AGTGTCTGTCACACTGTTTAAATTCGATGCTGGCCCATTAATCTCTGTACATAAAAGC	1205
DB	1141	AGTGTCTGTCACACTGTTTAAATTCGATGCTGGCCCATTAATCTCTGTACATAAAAGC	1200
QY	1206	TAGCAATGCGCTCTGGTATCAAAATGACTCCATTTGATGATCTACAGTGATTTAGATC	1265
DB	1201	TAGCAATGCGCTCTGGTATCAAAATGACTCCATTTGATGATCTACAGTGATTTAGATC	1260
QY	1266	GGTACTCAGCCAAACAGGCTATGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1325
DB	1261	GGTACTCAGCCAAACAGGCTATGCTCTTTTATATCAGGTCCCATGATGTGAAAAATGG	1320
QY	1326	AGGTGAATTTACTCATCCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG	1385
DB	1321	AGGTGAATTTACTCATCCCATAGCCCGGCGAGTCTCTCCCGCCCGCTCATCAG	1380
QY	1386	TCAGCGGTTGTCACCAACAAAGGCTGCGCCAGGCTTTATCGAACACAGCTTCCCTC	1445
DB	1381	TCAGCGGTTGTCACCAACAAAGGCTGCGCCAGGCTTTATCGAACACAGCTTCCCTC	1440
QY	1446	TCACATGATAAAGAAATCCACTCACTTAAATGGGACCTGGACATTTGAAGACACGCCAG	1505
DB	1441	TCACATGATAAAGAAATCCACTCACTTAAATGGGACCTGGACATTTGAAGACACGCCAG	1500

1506	QY	CAGTTCCTCATGTCGAGTCCCTAACCGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTTAATGC	1565
1501	Db		
1501	QY	CAGTTCCTCATGTCGAGTCCCTAACCGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTTAATGC	1560
1566	QY	TTTCAGCTTCTGTCCAAAACCTGCTCAGTTTAATAAGGTCCTCAGTGATCCCGAGAACATCCTTAA	1625
1561	Db		
1561	QY	TTTCAGCTTCTGTCCAAAACCTGCTCAGTTTAATAAGGTCCTCAGTGATCCCGAGAACATCCTTAA	1620
1626	QY	GAACAACAAAAATTACAATCAGTATTCAACAAGTTGCTGTTCCGACAGTGTCAAGTCCTCA	1685
1621	Db		
1621	QY	GAACAACAAAAATTACAATCAGTATTCAACAAGTTGCTGTTCCGACAGTGTCAAGTCCTCA	1680
1686	QY	ACCTAACCTTCATAGTAAATCTTTTGAGAACCCCTACCAAGCCCGTTTCCCTCTTCTTACCAT	1745
1681	Db		
1681	QY	ACCTAACCTTCATAGTAAATCTTTTGAGAACCCCTACCAAGCCCGTTTCCCTCTTCTTACCAT	1740
1746	QY	TACCAATTCGACGTACAGTCTACCTCGAACGCATCTACGATGTCCAGTTTCTAGTAAAGT	1805
1741	Db		
1741	QY	TACCAATTCGACGTACAGTCTACCTCGAACGCATCTACGATGTCCAGTTTCTAGTAAAGT	1800
1806	QY	AACAAAACCGATCCCCCGCAGTGAATCTGCTCCAGCCCGTGATGAATGGCAAAATCCAA	1865
1801	Db		
1801	QY	AACAAAACCGATCCCCCGCAGTGAATCTGCTCCAGCCCGTGATGAATGGCAAAATCCAA	1860
1866	QY	GCTGAACTCCAGCGTGTGCTGTCCTATGCGCGCAGTCTCTGAGGACTCTGACGAGA	1925
1861	Db		
1861	QY	GCTGAACTCCAGCGTGTGCTGTCCTATGCGCGCAGTCTCTGAGGACTCTGACGAGA	1920
1926	QY	GTCAAAGGGGCTGGGCAAGGAGAAATGGGATGGTACGATGTGAGTCTCCCACTCTCCCGG	1985
1921	Db		
1921	QY	GTCAAAGGGGCTGGGCAAGGAGAAATGGGATGGTACGATGTGAGTCTCCCACTCTCCCGG	1980
1986	QY	CCAAGATGCCGAAGATGAGGAGGCCATCTCCGACGAGCTTCAAGAACCCACGACCCCTTAA	2045
1981	Db		
1981	QY	CCAAGATGCCGAAGATGAGGAGGCCATCTCCGACGAGCTTCAAGAACCCACGACCCCTTAA	2040
2046	QY	CGGTGCTAATAGTGCAGACGACAGTACCGACGAGCTTCAAGAACCCACGACCCCTTAA	2105
2041	Db		
2041	QY	CGGTGCTAATAGTGCAGACGACAGTACCGACGAGCTTCAAGAACCCACGACCCCTTAA	2100
2106	QY	TGGCAGCTGCCAAGGCCAGCTGCTCCCTGCACTCAGAAAATCCCTTTGCTAAGGCAACCGG	2165
2101	Db		
2101	QY	TGGCAGCTGCCAAGGCCAGCTGCTCCCTGCACTCAGAAAATCCCTTTGCTAAGGCAACCGG	2160
2166	QY	TCTTCTCGGAAAGTTGATGCTGCTCTTCTGCTGTCTCTCCGAGACACAAATCTTAGA	2225
2161	Db		
2161	QY	TCTTCTCGGAAAGTTGATGCTGCTCTTCTGCTGTCTCTCCGAGACACAAATCTTAGA	2220
2226	QY	GACCTTCAGGCTTAGCAACAACTGAAAGGCTCCGACGATGAATGAGTGCACCTGGAGC	2285
2221	Db		
2221	QY	GACCTTCAGGCTTAGCAACAACTGAAAGGCTCCGACGATGAATGAGTGCACCTGGAGC	2280
2286	QY	AGAGAGGGGCGCTCCGAGGACCGCGACGCGAGCTCAGCTGGCAGCCCGCGCGCGA	2345
2281	Db		
2281	QY	AGAGAGGGGCGCTCCGAGGACCGCGACGCGAGCTCAGCTGGCAGCCCGCGCGCGA	2340
2346	QY	ATCCCTGGAGGAGCCAGATGCGGCGC	2372
2341	Db		
2341	QY	ATCCCTGGAGGAGCCAGATGCGGCGC	2367
RESULT 7			
BC060846			
LOCUS			
DEFINITION			
Homo sapiens ubiquitin specific protease 42, mRNA (cDNA clone			
IMAGE:30343487), partial cds.			
ACCESSION			
BC060846			
VERSION			
BC060846.1			
GI:38173811			
KEYWORDS			
SOURCE			
Homo sapiens (human)			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
BC060846			
3669 bp			
mRNA			
linear			
PRI 12-NOV-2003			

/translation="MDAGSASWAGVSSINDVSNHTLSILGPVPGAVVYSSSSVDPKSP SPQDQADGCIAPPQKVPSEKICLKWOOTHEVAGLQNLNWTFCANALQCLTIT PLIANYMLSEHSKTCHEGFCWMTQAHITQALSNPDVIFKPMFVINEMRRIRARF REGNEDAREFLQYTVDMQACLNKDRHTQATTLVQCIFGGYLSRVRCLNCK GVSDITFDYLDITLIEIKAAQSVNKAQFVKPQOLDGENSYKGSCKKMPVASKRFTI HRSNVLTLILKRFANFTGKIADVKYPIYLDITRPMYSQPNGEPIVIVLYAVLVHTG FNCHAGHYCYIKASNLGWTQMNDSIVSTSDIRSVLSQOAVLFIYRSHDVKNAGELT HPTFSPQSSPRPVIQSVORVTNKAAPFETGPOLPSHMIKNIPPHLNGTGPCLTPSS MSSPNGSNVNRASPVNASASVONWSNRSVITPEHPKQKQIITISHNKLPRVQCQSQ PNLHNSLENPTEKPVPSSTITNSAVQSTNSASTMSVSKVTEKIPRSESQSPVMGK SKLNSVIVPYGAESDDESKGLGKENGIGITIVSSHSPGQDADEEATPHELOSP MTLNGANSDDSKENGLAPDQASQCGQFALHSENPFPAKANGLPKMLPAPLLSIP EDKILFTRLSNLKSTDEMSAPGAERGPEDPDAEPQSPAALESLEPDAAGLS STKAPPPRPDPPTAPKEGAEAMAVAPPEPPSAGEDIVGDTAPDPLCPGSLTGDA SPLSQDAKMIASGRDLSALAEAPGLSPAPARSBEPCPOLIIVHPSGDHADAQDP SQSLGAPEAAERPPAPVLDMAPEGHEDAEPSFERVEDAAAPKAPGPSPAKEKIGS LRKDVHYARSRRSGSPARESKSTEGHRRRRRTCPREDRQDRHAPHPHGH GDLKSPGRRSLRCSHHSRSGSVELDWVRHYHTEGERGWREKFPYDPRPRWDR YHDRYALVAARDWKEFHGGRHERAGLHRPHKHNRGRGCEPARERHRPSRPR AGAPHALAPDRFSDHRTALVAGDNCNLSDRFHEHENGSKRKRHDSVENSISHVEK KARRSQKDPLEBPKAKKHKKKKKK" 393 1289 misc_feature /gene="USP42" /note="UCH; Region: Ubiquitin carboxyl-terminal hydrolase" /db_xref="CDD:pfam00443"		Query Match Best Local Similarity 93.7%; Score 2259; DB 9; Length 3669; Matches 2309; Conservative 0; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;	
QY	63	GAGTTGAACAAATGACCATAGTTGACAAAAGCTTCTGAATCTTCAGACCCCATCAGCCCTATCA	122
DB	53	GAGTTGAACAAATGACCATAGTTGACAAAAGCTTCTGAATCTTCAGACCCCATCAGCCCTATCA	112
QY	123	GAATCAGCTGGCAGCTCCGAGCAGTCTCACTGGAGACATGGATGCGAGTTCTGCCAG	182
DB	113	GAATCAGCTGGCAGCTCCGAGCAGTCTCACTGGAGACATGGATGCGAGTTCTGCCAG	172
QY	183	CTGGGTGCTGTCTTCAATGAATGATGTGCAATATCACTTTCTTTAGGACCAAT	242
DB	173	CTGGGTGCTGTCTTCAATGAATGATGTGCAATATCACTTTCTTTAGGACCAAT	232
QY	243	ACCTGGTGTGTGATTTATTCGAGTTTCATCTGTACTGTACTGATAAATCAAAACCATCACCA	302
DB	233	ACCTGGTGTGTGATTTATTCGAGTTTCATCTGTACTGTATAATCAAAACCATCACCA	292
QY	303	AAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTTCCCATCTGA	362
DB	293	AAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTTCCCATCTGA	352
QY	363	GAAGATTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGG	422
DB	353	GAAGATTGTCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGG	412
QY	423	CAATACCTGTTTGGCAATGACGACTGAGTGTGTTAACTACACACCACTCTTGGCCAA	482
DB	413	CAATACCTGTTTGGCAATGACGACTGAGTGTGTTAACTACACACCACTCTTGGCCAA	472
QY	483	TTACATGCTATCAGATGAACACTCCAAAACATGTCTATGAGAGGCTTTTGTATGATGTG	542
DB	473	TTACATGCTATCAGATGAACACTCCAAAACATGTCTATGAGAGGCTTTTGTATGATGTG	532
QY	543	TACATGAAGCACATATTAACCGGCACTAGTAATCTCTGGGACGTTTATTAACCAAT	602
DB	533	TACATGAAGCACATATTAACCGGCACTAGTAATCTCTGGGACGTTTATTAACCAAT	592
QY	603	GTTTGTATCAATGAGATCGCGCGTATAGCTAGGCACCTCGGTTTTTGAACCCAGAGA	662
DB	593	GTTTGTATCAATGAGATCGCGCGTATAGCTAGGCACCTCGGTTTTTGAACCCAGAGA	652
QY	663	TGCCCATGAATTCCTTCAATACACTGTTGATGCTATGAGAAAGCATGTTGAATGGCAG	722

DB	653	TGCCCATGAATTCCTTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAG	712
QY	723	CAATAAATTAGACAGACACACCCAGGGCCACACACTCTTTGTTGTCAGATATTTGGAGGATA	782
DB	713	CAATAAATTAGACAGACACACCCAGGGCCACACACTCTTTGTTGTCAGATATTTGGAGGATA	772
QY	783	CCTAAGACTCTAGAGTCAAAATGTTTAAATGCAAGGGCGTTTTCAGATATCTTTGATCCATA	842
DB	773	CCTAAGACTCTAGAGTCAAAATGTTTAAATGCAAGGGCGTTTTCAGATATCTTTGATCCATA	832
QY	843	TCCTTGATATAAATTGAGATAAAGCTGCTCAGAGTGTCAAAGGCAATGGAGAGATT	902
DB	833	TCCTTGATATAAATTGAGATAAAGCTGCTCAGAGTGTCAAAGGCAATGGAGAGATT	892
QY	903	TGTGAAGCCGGAACAGCTTCATGGAGAAAACTCGTCAAGTGCAGCAAGTGTAAAAAGAT	962
DB	893	TGTGAAGCCGGAACAGCTTCATGGAGAAAACTCGTCAAGTGCAGCAAGTGTAAAAAGAT	952
QY	963	GGTTCAGCTTCAAAGAGGTTCACTATCCATAGATCCCTAAATGTTCTTACACTTTCTCT	1022
DB	953	GGTTCAGCTTCAAAGAGGTTCACTATCCATAGATCCCTAAATGTTCTTACACTTTCTCT	1012
QY	1023	GAAACGTTTTTGCAAAATTTACCGGTGGAATAATGCTAAGGATGTGAAATACCTTGAGTA	1082
DB	1013	GAAACGTTTTTGCAAAATTTACCGGTGGAATAATGCTAAGGATGTGAAATACCTTGAGTA	1072
QY	1083	TCCTTGATATTCGGCCCATATATGTCCTCAACCCAGGAGAGCCCAATGTCCTACGTTCTTGT	1142
DB	1073	TCCTTGATATTCGGCCCATATATGTCCTCAACCCAGGAGAGCCCAATGTCCTACGTTCTTGT	1132
QY	1143	TGCAGTGTCTGGTCCACACACTGGTTTTAAATGGCCATGCTGGCCATTCTTCTGCTACATAA	1202
DB	1133	TGCAGTGTCTGGTCCACACACTGGTTTTAAATGGCCATGCTGGCCATTCTTCTGCTACATAA	1192
QY	1203	AGCTAGCAATGCGCTCTGCTGATCAATGAATGATGATGATGATGATGATGATGATGATGATG	1262
DB	1193	AGCTAGCAATGCGCTCTGCTGATCAATGAATGATGATGATGATGATGATGATGATGATGATG	1252
QY	1263	ATCGGTACTCAGCAACAAAGCCCTATGCTCTTTTATATATCAGGTCCCATGATGTGAAAAA	1322
DB	1253	ATCGGTACTCAGCAACAAAGCCCTATGCTCTTTTATATATCAGGTCCCATGATGTGAAAAA	1312
QY	1323	TGGAGTGAACCTACTCATCCACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCAT	1382
DB	1313	TGGAGTGAACCTACTCATCCACCATAGCCCGGCCAGTCTCTCCCGCCCGCTCAT	1372
QY	1383	CAGTCAGCGGGTGTCCACCAACAAACAGCTCGGCCAGGCTTTATCGGACCAACAGCTTCC	1442
DB	1373	CAGTCAGCGGGTGTCCACCAACAAACAGCTCGGCCAGGCTTTATCGGACCAACAGCTTCC	1432
QY	1443	CTCTCAGATGATAAAGAAATCCACCTCAGTTAAATGGGACTGGACCATTTGAAAGACACGCC	1502
DB	1433	CTCTCAGATGATAAAGAAATCCACCTCAGTTAAATGGGACTGGACCATTTGAAAGACACGCC	1492
QY	1503	AAGCAGTTCCATGTCGAGTCTTAACGGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTTAA	1562
DB	1493	AAGCAGTTCCATGTCGAGTCTTAACGGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTTAA	1552
QY	1563	TGCTTCAGCTTCTGTCCAAAACCTGGTCAGTTAAATAGGTCCTCAGTGATCCAGAACATCC	1622
DB	1553	TGCTTCAGCTTCTGTCCAAAACCTGGTCAGTTAAATAGGTCCTCAGTGATCCAGAACATCC	1612
QY	1623	TAAGAAAACAAAAATTAACAATCAGTATTCACAAAGTGTGCTGTTCCGCCAGTGTGAGTC	1682
DB	1613	TAAGAAAACAAAAATTAACAATCAGTATTCACAAAGTGTGCTGTTCCGCCAGTGTGAGTC	1672
QY	1683	TCAACCTTAACCTTCATAGTAATCTTTGGAGAAACCTTACCAAGCCCGTTCCTCTCTTAC	1742
DB	1673	TCAACCTTAACCTTCATAGTAATCTTTGGAGAAACCTTACCAAGCCCGTTCCTCTCTTAC	1732
QY	1743	CATTACCAATTCGCGAGTACAGTCTACTCGAAAGCATCTACGATGTGAGTCTTCTAGTAA	1802
DB	1733	CATTACCAATTCGCGAGTACAGTCTACTCGAAAGCATCTACGATGTGAGTCTTCTAGTAA	1792

Wed Aug 18 13:52:58 2004

QY	1803	AGTAACAAACCGATCCCGCCAGTGAATCTCTGCTCCCGAGCCGCTGATGAATGGCAATTC	1862
Db	1793	AGTAACAAACCGATCCCGCCAGTGAATCTCTGCTCCCGAGCCGCTGATGAATGGCAATTC	1852
QY	1863	CAAGCTGAATCTCAGCGTGTGTCCTATGGCGCGAGTCTCTGAGGACTCTCAACA	1922
Db	1853	CAAGCTGAATCTCAGCGTGTGTCCTATGGCGCGAGTCTCTGAGGACTCTCAACA	1912
QY	1923	GGAGTCAAAAGGGCTGGCAAGGAGATGGATTTGGTACGATTGTGAGCTCCCACTCTCC	1982
Db	1913	GGAGTCAAAAGGGCTGGCAAGGAGATGGATTTGGTACGATTGTGAGCTCCCACTCTCC	1972
QY	1983	CGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCT	2042
Db	1973	CGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCT	2032
QY	2043	AAACGGTGTCTAATAGTGCAGACAGACAGTGAACCTGACCGAAGAAAGCGCTAGCGCTGA	2102
Db	2033	AAACGGTGTCTAATAGTGCAGACAGACAGTGAACCTGACCGAAGAAAGCGCTAGCGCTGA	2092
QY	2103	TGGTGCAGCTGCAAGGGCAGCGCTGCGCTGCACTCAGAAAAATCCCTTTTCTAAGGCAAA	2162
Db	2093	TGGTGCAGCTGCAAGGGCAGCGCTGCGCTGCACTCAGAAAAATCCCTTTTCTAAGGCAAA	2152
QY	2163	CGGTCTTCTCGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2222
Db	2153	CGGTCTTCTCGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2212
QY	2223	AGAGACCTTCAGCTTAGCAACAACTGAAGGCTCGACGAGTGAATGAGTGCACCTGG	2282
Db	2213	AGAGACCTTCAGCTTAGCAACAACTGAAGGCTCGACGAGTGAATGAGTGCACCTGG	2272
QY	2283	AGCAGAGAGGGCCCTCCGAGGACCGGACCGAGCGCTCAGCGCTCGAGCGCCCGCCGCG	2342
Db	2273	AGCAGAGAGGGCCCTCCGAGGACCGGACCGAGCGCTCAGCGCTCGAGCGCCCGCCGCG	2332
QY	2343	CGAATCCCTGGAGGAGCAGATGGCGCGC	2372
Db	2333	CGAATCCCTGGAGGAGCAGATGGCGCGC	2362
RESULT 8			
AX364931			
LOCUS			
Sequence 82 from Patent WO0206315.			
AX364931			
AX364931.1 GI:18696821			
KEYWORDS			
SOURCE			
ORGANISM			
Homo sapiens (human)			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
Mintz, L., Freilich, S. and Bernstein, J.			
Novel nucleic acid and amino acid sequences			
Patent: WO 0206315-A 82 24-JAN-2002;			
CompuGen Ltd. (IL)			
FEATURES			
Location/Qualifiers			
1..1679			
/organism="Homo sapiens"			
/mol_type="unassigned DNA"			
/db_xref="taxon:9606"			
ORIGIN			
Query Match 61.2%; Score 1475; DB 6; Length 1679;			
Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1675; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	694	GCTATGCAAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACC	753
Db	1	GCTATGCAAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACC	60

QY	754	ACTCTTGTTGTGAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTC	813
Db	61	ACTCTTGTTGTGAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTC	120
QY	814	AAGGGCGTTTCAGATACCTTTTGTATCATATCTTTGATATTAACATTTGAGATAAAGGCTGCT	873
Db	121	AAGGGCGTTTCAGATACCTTTTGTATCATATCTTTGATATTAACATTTGAGATAAAGGCTGCT	180
QY	874	CAGAGTGTCAAACAGCATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGAGAGAAAAC	933
Db	181	CAGAGTGTCAAACAGCATTTGGAGCAGTTTGTGAAGCCGGAACAGCTTGATGAGAGAAAAC	240
QY	934	TGCTACAAGTGCAGCAAGTGTAAAAAGATGGTTCCAGCTTCAAAGAGTTTCACTATCCAT	993
Db	241	TGCTACAAGTGCAGCAAGTGTAAAAAGATGGTTCCAGCTTCAAAGAGTTTCACTATCCAT	300
QY	994	AGATCCCTCTAATGTTCTTACACTTTCTGTAACGTTTTCGCAAAATTTTACCGGTGGAAA	1053
Db	301	AGATCCCTCTAATGTTCTTACACTTTCTGTAACGTTTTCGCAAAATTTTACCGGTGGAAA	360
QY	1054	ATTGCTAAGGATGTGAATAACCTGAGTATCTTTGATTCGGCCATATATGTCTCAACCC	1113
Db	361	ATTGCTAAGGATGTGAATAACCTGAGTATCTTTGATTCGGCCATATATGTCTCAACCC	420
QY	1114	AACGGAGAGCAATTTGTCTACGTTCTTGTATGCAAGTGTGTCCACACTGTTTAAATTC	1173
Db	421	AACGGAGAGCAATTTGTCTACGTTCTTGTATGCAAGTGTGTCCACACTGTTTAAATTC	480
QY	1174	CATGCTGGCCATTTACTTCTGCTACATAAAGCTTAGCAATGSCCTCTGATCAATGAAT	1233
Db	481	CATGCTGGCCATTTACTTCTGCTACATAAAGCTTAGCAATGSCCTCTGATCAATGAAT	540
QY	1234	GACTCCATTTGATCTACCAAGTATATTAGATCGGTACTCAGCCCAACAGCCCTATGTGCTC	1293
Db	541	GACTCCATTTGATCTACCAAGTATATTAGATCGGTACTCAGCCCAACAGCCCTATGTGCTC	600
QY	1294	TTTTATATCAGGTCCTCATGATGTGAATAATGGAGTGAACTTACTCATCCACCCTATG	1353
Db	601	TTTTATATCAGGTCCTCATGATGTGAATAATGGAGTGAACTTACTCATCCACCCTATG	660
QY	1354	CCCGGCCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGCACCAACAAACAGCT	1413
Db	661	CCCGGCCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGGTGTGCACCAACAAACAGCT	720
QY	1414	GGCGCCAGCTTTATCGGACCAACAGCTTCCCTCTCACATGATAAAGAAATCCACCTC	1473
Db	721	GGCGCCAGCTTTATCGGACCAACAGCTTCCCTCTCACATGATAAAGAAATCCACCTC	780
QY	1474	AATGGGACTGGACCAATTTGAAAGACACGCAAGCAGTTCCATGTGAGTCTTAAACGGG	1533
Db	781	AATGGGACTGGACCAATTTGAAAGACACGCAAGCAGTTCCATGTGAGTCTTAAACGGG	840
QY	1534	TCCAGTGTCAACAGAGCTAGTCTCTGTTAAATGCTTTCAGCTTCTGTCACAAATCGT	1593
Db	841	TCCAGTGTCAACAGAGCTAGTCTCTGTTAAATGCTTTCAGCTTCTGTCACAAATCGT	900
QY	1594	AATAGTCTCTCAGTATCCAGAACATCCTTAAGAAACAAAAAATTAACATCAGTATTC	1653
Db	901	AATAGTCTCTCAGTATCCAGAACATCCTTAAGAAACAAAAAATTAACATCAGTATTC	960
QY	1654	AACAGTGTGCTGTTTCGGCCAGTGTGAGTCTCAACCTAACCTTCACTAGTAAATCTT	1713
Db	961	AACAGTGTGCTGTTTCGGCCAGTGTGAGTCTCAACCTAACCTTCACTAGTAAATCTT	1020
QY	1714	AACCTTACCAAGCCCGTTCCCTCTTCTACCAATTTCTGCAAGTACAGTCTACCTCG	1773
Db	1021	AACCTTACCAAGCCCGTTCCCTCTTCTACCAATTTCTGCAAGTACAGTCTACCTCG	1080
QY	1774	AACGATCTACGATCTCAGTCTTCTAGTAAAGTAAACAAACCGATCCCGCGAGTGAATCC	1833
Db	1081	AACGATCTACGATCTCAGTCTTCTAGTAAAGTAAACAAACCGATCCCGCGAGTGAATCC	1140
QY	1834	TGCTCCAGCCCGCTGATGAAATGGCAAAATCCAAAGCTGAATCCAGCGTGTGGTCCCTAT	1893

1141	TGCTCCAGCCCGTGATGAATGGCAAAATCAAGCTGAATCCAGCGTCTGGTGCCCTAT	1200
1894	GGCCCGGAGTCCTCTGAGGACTCTGACGAGAGTCAAAAGGGCTGGCGCAGGAGAATGGG	1953
1201	GGCCCGGAGTCCTCTGAGGACTCTGACGAGAGTCAAAAGGGCTGGCGCAGGAGAATGGG	1260
1954	ATTGGTAGGATTGTGAGCTCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACT	2013
1261	ATTGGTAGGATTGTGAGCTCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACT	1320
2014	CCGCAAGAGCTTCAAAGAAACCATGACCCTAAACGGTCTAATAGTCAGACACGGACAGT	2073
1321	CCGCAAGAGCTTCAAAGAAACCATGACCCTAATAGTCAGACACGGACAGT	1380
2074	GACCCGAAAGAAACGGCCTTAGCGCTGTATGTCGACAGTCCCAAGGCCAGCTGCCTCG	2133
1381	GACCCGAAAGAAACGGCCTTAGCGCTGTATGTCGACAGTCCCAAGGCCAGCTGCCTCG	1440
2134	CACTCAGAAATCCCTTTGTCTAAGGCAACCGGTCTTCTCGAAAGTTGATGCTGCTCCT	2193
1441	CACTCAGAAATCCCTTTGTCTAAGGCAACCGGTCTTCTCGAAAGTTGATGCTGCTCCT	1500
2194	TTGCTGCTCTCCGAGNAGCAAAATCTTAGACCTTCAGGCTTAGACACAAACTGAAA	2253
1501	TTGCTGCTCTCCGAGNAGCAAAATCTTAGACCTTCAGGCTTAGACACAAACTGAAA	1560
2254	GGCTCGACGATGAATAGTGTGACCTCGAGCAGAGAGGGGCCCTCCCGAGACCCGAC	2313
1561	GGCTCGACGATGAATAGTGTGACCTCGAGCAGAGAGGGGCCCTCCCGAGACCCGAC	1620
2314	GCCGAGCCTCAGCCTGGCAGCCCGCCGCGAATCTCTGAGAGCCAGATCGGCCGC	2372
1621	GCCGAGCCTCAGCCTGGCAGCCCGCCGCGAATCTCTGAGAGCCAGATCGGCCGC	1679

RESULT 9

AC004895/c	AC004895	152927 bp	DNA	linear	PRI 30-SEP-2000
LOCUS	Hom sapiens PAC clone RP4-810E6				
DEFINITION	AC004895				
ACCESSION	AC004895.2	GI:4926908			
VERSION	HTG.				
KEYWORDS					
SOURCE	Hom sapiens (human)				
ORGANISM	Hom sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 152927)				
AUTHORS	Sulston,J.E. and Waterston,R.				
TITLE	Toward a complete human genome sequence				
JOURNAL	Genome Res. 8 (11), 1097-1108 (1998)				
MEDLINE	99063792				
PUBMED	9847074				
REFERENCE	2 (bases 1 to 152927)				
AUTHORS	Cordes,M. and Gibson,A.				
TITLE	The sequence of Homo sapiens PAC clone RP4-810E6				
JOURNAL	Unpublished				
REFERENCE	3 (bases 1 to 152927)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-JUN-1998) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	4 (bases 1 to 152927)				
AUTHORS	Waterston,R.H.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA				
REFERENCE	5 (bases 1 to 152927)				
AUTHORS	Waterston,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-SEP-2000) Department of Genetics, Washington				

repeat_region	/rpt_family="Alu"	2977..3093
repeat_region	/rpt_family="L1"	3094..3392
repeat_region	/rpt_family="Alu"	3393..3448
repeat_region	/rpt_family="L1"	3449..3751
repeat_region	/rpt_family="Alu"	3761..4060
repeat_region	/rpt_family="Alu"	4061..4162
repeat_region	/rpt_family="L1"	4163..4485
repeat_region	/rpt_family="Alu"	4486..4573
repeat_region	/rpt_family="L1"	4574..4689
repeat_region	/rpt_family="Alu"	4691..4888
repeat_region	/rpt_family="L1"	4969..4988
repeat_region	/rpt_family=" (TAAA)n"	4992..5112
repeat_region	/rpt_family="Alu"	5121..5179
repeat_region	/rpt_family="L1"	5180..5486
repeat_region	/rpt_family="Alu"	5487..5736
repeat_region	/rpt_family="L1"	5912..5965
repeat_region	/rpt_family="A-rich"	6214..6526
repeat_region	/rpt_family="Alu"	6544..6601
repeat_region	/rpt_family="L2"	6673..6984
repeat_region	/rpt_family="Alu"	7523..7625
repeat_region	/rpt_family="MIR"	7914..8236
repeat_region	/rpt_family="Alu"	8558..8624
misc_feature	/note="match to EST M1"	9165..9231
repeat_region	/note="match to EST M1"	10016..10318
repeat_region	/rpt_family="Alu"	10529..10564
repeat_region	/rpt_family="Alu"	10565..10867
repeat_region	/rpt_family="Alu"	10868..10977
repeat_region	/rpt_family="Alu"	11402..11705
repeat_region	/rpt_family="Alu"	11922..12231
repeat_region	/rpt_family="Alu"	12233..12567
repeat_region	/rpt_family="MaLR"	13482..13754
repeat_region	/rpt_family="Alu"	13771..13925
repeat_region	/rpt_family="Alu"	13932..14240
repeat_region	/rpt_family="Alu"	14520..14937
repeat_region	/rpt_family="Other"	15405..15431
repeat_region	/rpt_family="AT-rich"	15611..15852
repeat_region	/rpt_family="MaLR"	

repeat_region	16336..16460	/rpt_family="(CA)n"
repeat_region	16480..16693	/rpt_family="(CA)n"
repeat_region	16793..17187	/rpt_family="(L1)"
repeat_region	17675..17877	/rpt_family="(MaLR)"
misc_feature	18340..18452	/note="match to EST M79246 (NID:g273559)"
misc_feature	18400..18458	/note="similar to EST AA284549 (NID:g1927478) zt24f01.r1"
repeat_region	19073..19720	/rpt_family="(L1)"
repeat_region	19721..20026	/rpt_family="(Alu)"
repeat_region	20027..20435	/rpt_family="(L1)"
repeat_region	20436..20735	
Query Match 29.9%; Score 722; DB 9; Length 152927;		
Best Local Similarity 100.0%; Pred. No. 0;		
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	1457	AGATCCACTCCTTAATGGGACTGGACCTTGAAGAGACCGCCAGCAGTTCCTCATGT 1516
Db	46699	AGAATCCACTCCTTAATGGGACTGGACCTTGAAGAGACCGCCAGCAGTTCCTCATGT 46640
QY	1517	CGAGTCCTTAACGGGATTCAGCTGTCAACAGGGCTAGTCCTGTAAATGCTTCAGCTTCTG 1576
Db	46639	CGAGTCCTTAACGGGATTCAGCTGTCAACAGGGCTAGTCCTGTAAATGCTTCAGCTTCTG 46580
QY	1577	TCCAAACTGGTTCAGTTAATAGTCTCTCAGTGATCCCAGAACATCCTAAGAAACAAAAA 1636
Db	46579	TCCAAACTGGTTCAGTTAATAGTCTCTCAGTGATCCCAGAACATCCTAAGAAACAAAAA 46520
QY	1637	TTCAATTCAGTATTTCAACAAGTTGCTGTTCGCCAGTGTTCAGTCTCAACTAACCTTC 1696
Db	46519	TTCAATTCAGTATTTCAACAAGTTGCTGTTCGCCAGTGTTCAGTCTCAACTAACCTTC 46460
QY	1697	ATAGTAACTTTTGGAGAACCTTACCAGCCGTTCCCTCTTCTTACCATTACCAATTCGT 1756
Db	46459	ATAGTAACTTTTGGAGAACCTTACCAGCCGTTCCCTCTTCTTACCATTACCAATTCGT 46400
QY	1757	CAGTACAGTCTACTCGAAGCGCATTCAGATGTCAAGTTCTAGTAAAGTAAACCAACCGA 1816
Db	46399	CAGTACAGTCTACTCGAAGCGCATTCAGATGTCAAGTTCTAGTAAAGTAAACCAACCGA 46340
QY	1817	TCCCCCGCAGTGAACTCTGTCTCCAGCCGTCATGAATGGCAAAATCCAGCTGAATCCA 1876
Db	46339	TCCCCCGCAGTGAACTCTGTCTCCAGCCGTCATGAATGGCAAAATCCAGCTGAATCCA 46280
QY	1877	CGCTGCTGGTGCCTATGGCCGCGAGTCTCTTGAGACTCTCAGCAGGAGTCAAAAGGGGC 1936
Db	46279	CGCTGCTGGTGCCTATGGCCGCGAGTCTCTTGAGACTCTCAGCAGGAGTCAAAAGGGGC 46220
QY	1937	TGGCAAGAGAGATGGATTGGTACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCG 1996
Db	46219	TGGCAAGAGAGATGGATTGGTACGATTGTGAGCTCTCCCGGCCAAGATGCCG 46160
QY	1997	AAGATGAGAGGCCACTCTCCGACAGCTTCAAGAACCCATGACCTTAAACGGTGTAAATA 2056
Db	46159	AAGATGAGAGGCCACTCTCCGACAGCTTCAAGAACCCATGACCTTAAACGGTGTAAATA 46100
QY	2057	GTGCAGACGACGACGATGACCCGAAAGAAACGGCTAGCGCTGTATGGTGCAGCTGCC 2116
Db	46099	GTGCAGACGACGACGATGACCCGAAAGAAACGGCTAGCGCTGTATGGTGCAGCTGCC 46040
QY	2117	AAGGCCAGCTTGCCTGTCAGCTCAGAAAAATCCCTTTGCTAAGGCAACGGTTCCTGGAA 2176
Db	46039	AAGGCCAGCTTGCCTGTCAGCTCAGAAAAATCCCTTTGCTAAGGCAACGGTTCCTGGAA 45980
QY	2177	AG 2178


```

Db      45979 AG 45978

RESULT 10
AC069393
LOCUS   154791 bp DNA linear HTG 26-OCT-2000
DEFINITION Homo sapiens chromosome 7 clone RP11-585K18, WORKING DRAFT
AC069393
ACCESSION AC069393.8 GI:11024947
VERSION   HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS  Homo sapiens (human)
SOURCE    Homo sapiens
ORGANISM  Homo sapiens
          Bunkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
          Waterston,R.H.
          The sequence of Homo sapiens clone
REFERENCE Unpublished
AUTHORS   2 (bases 1 to 154791)
TITLE     Waterston,R.H.
JOURNAL   Direct Submission
AUTHORS   Submitted (28-MAY-2000) Genome Sequencing Center, Washington
TITLE     University School of Medicine, 4444 Forest Park Parkway, St. Louis,
SOURCE    MO 63108, USA
COMMENT   On Oct 26, 2000 this sequence version replaced gi:10946568.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0585K18
----- Summary Statistics -----
Sequencing vector: M13; 71%
Sequencing vector: plasmid; 5%
Chemistry: Dye-primer ET; 71% of reads
Chemistry: Dye-terminator Big Dye; 5% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 143345 bases at least Q40
Consensus quality: 146572 bases at least Q30
Consensus quality: 148386 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 153491; sum-of-contigs
Quality coverage: 4.65 in Q20 bases; agarose-fp
Quality coverage: 5.04 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 5209: contig of 5209 bp in length
5210 5309: gap of unknown length
5310 10071: contig of 4762 bp in length
10072 10171: gap of unknown length
10172 16164: contig of 5993 bp in length
16165 16264: gap of unknown length
16265 29394: contig of 13130 bp in length
29395 29494: gap of unknown length
29495 41590: contig of 12096 bp in length
41591 41690: gap of unknown length
41691 54703: contig of 13013 bp in length
54704 54803: gap of unknown length
54804 74381: contig of 19578 bp in length
74382 74481: gap of unknown length
103253 103254: contig of 28772 bp in length
103254 103354: gap of unknown length
103354 139431: contig of 36078 bp in length

```

```

* 139432 139531: gap of unknown length
* 139532 140718: contig of 1187 bp in length
* 140719 140818: gap of unknown length
* 140819 143471: contig of 2653 bp in length
* 143472 143571: gap of unknown length
* 143572 147062: contig of 3491 bp in length
* 147063 147162: gap of unknown length
* 147163 150060: contig of 2898 bp in length
* 150061 150160: gap of unknown length
* 150161 154791: contig of 4631 bp in length.
FEATURES
        Location/Qualifiers
            1..154791
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="7"
            /clone="RP11-585K18"
            1..5209
            /note="assembly_name:Contig10"
            5310..10071
            /note="assembly_name:Contig11"
            10172..16164
            /note="assembly_name:Contig12"
            clone_end:T7
            vector_side:left
            16265..29394
            /note="assembly_name:Contig13"
            29495..41590
            /note="assembly_name:Contig14"
            41691..54703
            /note="assembly_name:Contig15"
            54804..74381
            /note="assembly_name:Contig16"
            74482..103253
            /note="assembly_name:Contig17"
            103354..139431
            /note="assembly_name:Contig18"
            139532..140718
            /note="assembly_name:Contig19"
            clone_end:SP6
            vector_side:right
            140819..143471
            /note="assembly_name:Contig6"
            143572..147062
            /note="assembly_name:Contig7"
            147163..150060
            /note="assembly_name:Contig8"
            150161..154791
            /note="assembly_name:Contig9"
ORIGIN
Query Match      29.9%; Score 722; DB 2; Length 154791;
Best Local Similarity 100.0%; Pred.No. 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1457 AGAATCCACTCTCACTTAATGGGACTGGACCATTAAGAGACACGCCAAGAGATTCCATGT 1516
Db 129201 AGAATCCACTCTCACTTAATGGGACTGGACCATTAAGAGACACGCCAAGAGATTCCATGT 129260
QY 1517 CGAGTCTCTAACGGGAATTCAGAGTGTCAACAGGGCTAGTCTCTGTTAATGCTCAGCTTCTG 1576
Db 129261 CGAGTCTCTAACGGGAATTCAGAGTGTCAACAGGGCTAGTCTCTGTTAATGCTCAGCTTCTG 129320
QY 1577 TCCAAAACCTGGTCAGTAAATAGGTCTCTCAGTGATCCAGAACATCTCTAAGAAACAAAAA 1636
Db 129321 TCCAAAACCTGGTCAGTAAATAGGTCTCTCAGTGATCCAGAACATCTCTAAGAAACAAAAA 129380
QY 1637 TTACAATCAGTATTCAACAAGTTGCTGTTTCCGCCAGTGTCTCAGTCTCAACCTTAACCTTC 1696
Db 129381 TTACAATCAGTATTCAACAAGTTGCTGTTTCCGCCAGTGTCTCAGTCTCAACCTTAACCTTC 129440
QY 1697 ATAGTAATCTTTGGAGAACCCCTCAACAGCCCGTTCCCTCTTCTACCAATTACCAATTCGT 1756

```

Consensus quality: 208289 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 26 contigs, the true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

Db	129441	ATAGTAATTCTTTGGAGAAACCTACCAAGCCCGTTCCCTCTTCTACCAATTACCAATCTG	129500
Qy	1757	CAGTACAGTCTACCTCGAAGCGCATCTACGATGTGTCAGTTCTAGTAAAGTAAACAAACCGA	1816
Db	129501	CAGTACAGTCTACCTCGAAGCGCATCTACGATGTGTCAGTTCTAGTAAAGTAAACAAACCGA	129560
Qy	1817	TCCCCCGCAGTGAATCTGCTCCAGCCCGTGATGATGGCAAAATCCAAAGCTGAATCCA	1876
Db	129561	TCCCCCGCAGTGAATCTGCTCCAGCCCGTGATGATGGCAAAATCCAAAGCTGAATCCA	129620
Qy	1877	GCGTGCTGCTCCCTATGCGCCCGAGTCTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGC	1936
Db	129621	GCGTGCTGCTCCCTATGCGCCCGAGTCTCTCTGAGGACTCTGACGAGGAGTCAAAAGGGC	129680
Qy	1937	TGGCACAAGGAGAATGGGATTTGATACGATTGTGAGCTCCCACTTCCCGCCCAAGATCGC	1996
Db	129681	TGGCACAAGGAGAATGGGATTTGATACGATTGTGAGCTCCCACTTCCCGCCCAAGATCGC	129740
Qy	1997	AAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCCATGACCCCTAAACGGTGTCTAATA	2056
Db	129741	AAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCCATGACCCCTAAACGGTGTCTAATA	129800
Qy	2057	GTGCACAGCAGCAGTGAATGATGACCCGAAAGAAACCGCTAGCGCTGATGTGCCAGCTGCC	2116
Db	129801	GTGCACAGCAGCAGTGAATGATGACCCGAAAGAAACCGCTAGCGCTGATGTGCCAGCTGCC	129860
Qy	2117	AAGGCCAGCCTGCCCTGCACTCAGAAATCCCTTTGCTTAAGCCAAACGGTCTTCCTGGAA	2176
Db	129861	AAGGCCAGCCTGCCCTGCACTCAGAAATCCCTTTGCTTAAGCCAAACGGTCTTCCTGGAA	129920
Qy	2177	AG 2178	
Db	129921	AG 129922	
RESULT 11			
AC146146/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			

DEFINITION Pan troglodytes chromosome UNK clone RP43-2G9, WORKING DRAFT
 SEQUENCE, 26 unordered pieces.
 AC146146
 VERSION AC146146.1 GI-33387100
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 1 (bases 1 to 217022)
 REFERENCE Wilson,R.K.
 AUTHORS The sequence of Pan troglodytes clone
 TITLE Unpublished
 JOURNAL
 REFERENCE 2 (bases 1 to 217022)
 AUTHORS Wilson,R.K.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444
 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@watson.wustl.edu
 ----- Project Information -----
 Center project name: C_PT002G09
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Chemistry: Dye-primer ET; 0% of reads
 Assembly: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 204681 bases at least Q40
 Consensus quality: 206693 bases at least Q30
 Consensus quality: 208289 bases at least Q20
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 26 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1098: contig of 1098 bp in length
 * 1099 1198: gap of unknown length
 * 1199 2982: contig of 1784 bp in length
 * 2983 3082: gap of unknown length
 * 3083 5180: contig of 2098 bp in length
 * 5181 5280: gap of unknown length
 * 5281 7152: contig of 1872 bp in length
 * 7153 7252: gap of unknown length
 * 7253 9449: contig of 2197 bp in length
 * 9450 9549: gap of unknown length
 * 9550 10943: contig of 1394 bp in length
 * 10944 11044: gap of unknown length
 * 11044 13812: contig of 2769 bp in length
 * 13813 13913: gap of unknown length
 * 13913 18353: contig of 4441 bp in length
 * 18354 18454: gap of unknown length
 * 18454 21675: contig of 3222 bp in length
 * 21676 21775: gap of unknown length
 * 21776 24307: contig of 2532 bp in length
 * 24308 24407: gap of unknown length
 * 24408 28212: contig of 3805 bp in length
 * 28213 34694: contig of 6382 bp in length
 * 34695 34794: gap of unknown length
 * 34795 38278: contig of 3484 bp in length
 * 38279 38378: gap of unknown length
 * 38379 44872: contig of 6494 bp in length
 * 44873 44972: gap of unknown length
 * 44973 51522: contig of 6550 bp in length
 * 51523 51622: gap of unknown length
 * 51623 59330: contig of 7708 bp in length
 * 59331 59430: gap of unknown length
 * 59431 67685: contig of 8255 bp in length
 * 67686 67785: gap of unknown length
 * 67786 76304: contig of 8519 bp in length
 * 76305 76404: gap of unknown length
 * 76405 91684: contig of 15280 bp in length
 * 91685 91784: gap of unknown length
 * 91785 103781: contig of 11997 bp in length
 * 103782 103882: gap of unknown length
 * 103883 115984: contig of 12102 bp in length
 * 115985 131828: contig of 15745 bp in length
 * 131829 131928: gap of unknown length
 * 131929 148774: contig of 16846 bp in length
 * 148775 148874: gap of unknown length
 * 148875 168325: contig of 19451 bp in length
 * 168326 168425: gap of unknown length
 * 168426 185595: contig of 17170 bp in length
 * 185596 217022: contig of 31327 bp in length.
 FEATURES
 Location/Qualifiers
 1..217022
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /chromosome="UNK"
 /clone="RP43-2G9"
 1..1098
 /note="assembly_name:Contig18"
 1199..2982
 /note="assembly_name:Contig23"
 3083..5180
 /note="assembly_name:Contig25"
 5281..7152
 /note="assembly_name:Contig26"
 7253..9449
 /note="assembly_name:Contig27"
 9550..10943
 /note="assembly_name:Contig28"
 11044..13812
 /note="assembly_name:Contig29"
 13913..18353
 /note="assembly_name:Contig30"
 18454..21675
 /note="assembly_name:Contig31"
 21776..24307
 /note="assembly_name:Contig32"
 24408..28212
 /note="assembly_name:Contig33"
 28313..34694
 /note="assembly_name:Contig34"
 34795..38278
 /note="assembly_name:Contig35"
 38379..44872
 /note="assembly_name:Contig36"
 44973..51522
 /note="assembly_name:Contig37"
 51623..59330
 /note="assembly_name:Contig38"
 59431..67685
 /note="assembly_name:Contig39"
 67786..76304
 /note="assembly_name:Contig40"
 76405..91684
 /note="assembly_name:Contig41"
 91785..103781
 /note="assembly_name:Contig42"
 103882..115983
 /note="assembly_name:Contig43"
 116084..131828
 /note="assembly_name:Contig44"
 131829..148774
 /note="assembly_name:Contig45"
 148775..168325
 /note="assembly_name:Contig46"
 168326..185595
 /note="assembly_name:Contig47"
 185596..217022
 /note="assembly_name:Contig48"

[illegible]

<hr/>					
BD147651					
BD147651.1 GI:27853409					
JP 2002191363-A/2494					
Homo sapiens (human)					
Homo sapiens					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
1 (bases 1 to 521)					
Ota,T., Isogai,T., Nishikawa,T., Hayashi,K., Saito,K., Yamamoto,J.,					
Ishii,S., Sugiyama,T., Wakamatsu,A., Nagai,K. and Otsuki,T.					
Primer for synthesizing full-length cDNA and use thereof					
Patent: JP 2002191363-A 2494 09-JUL-2002;					
HELIIX RESEARCH INSTITUTE					
OS Homo sapiens (human)					
PN JP 2002191363-A/2494					
PD 09-JUL-2002					
PF 28-JUL-2000 JP 2000280990					
PI TOSHIO OTA, TAKAO ISOGAI, TETSUO NISHIKAWA, KOJI HAYASHI, KAORU					
PI SAITO,					
PI JUNICHI YAMAMOTO, SHIZUKO ISHII, TOMOYASU SUGIYAMA, AI WAKAMATSU,					
PI KEIICHI NAGAI, TETSUOI OTSUKI					
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/ PC					
10, C12N15/02,C12N1/68/C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC					
Primer for synthesizing full-length cDNA and use thereof FH Key					
Location/Qualifiers					
FT source 1..521					
FT Location/Qualifiers					
1..521					
/organism='Homo sapiens'					
/organism='Homo sapiens'					
/mol_type='genomic DNA'					
/db_xref='taxon:9606'					
<hr/>					
ORIGIN					
Query Match 9.6%; Score 232; DB 6; Length 521;					
Best Local Similarity 99.6%; Pred No. 2.7e-123;					
Matches 282; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
<hr/>					
QY	6	GTGTCGGCGCGCGCGCGCGCGCGCGGATGGAGCGGCAGCGCGGCTCAGACCCATCAGCCTATCAGAA	125		
DB	1	GTGTCGGCGCGCGCGCGCGCGCGGATGGAGCGGCAGCGCGGCTCAGACCCATCAGCCTATCAGAA	60		
<hr/>					
QY	66	TGAACAATGACCATTAGTGAACAAGCTTCTCAATCTTCAGACCCATCAGCCTATCAGAA	125		
DB	61	TGAACAATGACCATTAGTGAACAAGCTTCTCAATCTTCAGACCCATCAGCCTATCAGAA	120		
<hr/>					
QY	126	TCAGCTGCGCAGCTCGGAGCGAGTCTCACCTGGAGCATGGATGACGTTTCGCCAGTG	185		
DB	121	TCAGCTGCGCAGCTCGGAGCGAGTCTCACCTGGAGCATGGATGACGTTTCGCCAGTG	180		
<hr/>					
QY	186	GCGTGCTGTGTCTTCATTAATAATGATGTCAAATCACACTTCTTTAGGACCGAGTACC	245		
DB	181	GCGTGCTGTGTCTTCATTAATAATGATGTCAAATCACACTTCTTTAGGACCGAGTACC	240		
<hr/>					
QY	246	TGGTGCTGTAGTTTATTCGAGTTTCATCTGTACCTGATAATCA	288		
DB	241	TGGTGCTGTAGTTTATTCGAGTTTCATCTGTACCTGATAATCA	283		
<hr/>					
RESULT 17					
AR264036					
LOCUS					
DEFINITION Sequence 214 from patent US 6331427.					
ACCESSION AR264036					
VERSION AR264036.1 GI:28076040					
KEYWORDS					
SOURCE Unknown.					
ORGANISM Unknown.					
REFERENCE 1 (bases 1 to 457)					
AUTHORS Robinson,K.E.					
PAT 29-JAN-2000					

TITLE		Protease homologs	
JOURNAL	Patent: US 633142;-A 214 18-DEC-2001;		
FEATURES	Location/Qualifiers		
source	1..457		
	/organism="unknown"		
	/mol_type="genomic DNA"		
ORIGIN			
Query Match	8.5%;	Score 205;	DB 6; Length 457;
Best Local Similarity	100.0%;	Pred. No. 1.8e-107;	Indels 0; Gaps 0;
Matches 205;	Conservative 0;	Mismatches 0;	
QY	138	CTCCGAGGAGTCTACCTGGAGACATGGATGCAGGTTCTGCCAGCTGGGGTGTGTGTC	197
Db	103	CTCCGAGGAGTCTACCTGGAGACATGGATGCAGGTTCTGCCAGCTGGGGTGTGTGTC	162
QY	198	TTCAATTGAATGATGTGTCAAATCACACACTTTCTTTAGGACAGTACCTGGTGTGTAGT	257
Db	163	TTCAATTGAATGATGTGTCAAATCACACACTTTCTTTAGGACAGTACCTGGTGTGTAGT	222
QY	258	TTATTCGAGTTCACTGTGTACCTGTATAATCAAAACCATCACACAAAAGGATCAAGCCCT	317
Db	223	TTATTCGAGTTCACTGTGTACCTGTATAATCAAAACCATCACACAAAAGGATCAAGCCCT	282
QY	318	AGGTGATGGCATCGCTCCTCCACAG	342
Db	283	AGGTGATGGCATCGCTCCTCCACAG	307
RESULT 18			
AC069393/c	AC069393 154791 bp DNA linear HTG 26-OCT-2000		
LOCUS	Homo sapiens chromosome 7 clone RP11-585K18, WORKING DRAFT		
DEFINITION	SEQUENCE, 14 unordered pieces.		
ACCESSION	AC069393		
VERSION	AC069393.8 GI:11024947		
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Waterston,R.H.		
TITLE	The sequence of Homo sapiens clone		
JOURNAL	Unpublished		
AUTHORS	Waterston,R.H.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-MAY-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA		
COMMENT	On Oct 26, 2000 this sequence version replaced gi:10946568.		
----- Genome Center -----			
Center: Washington University Genome Sequencing Center			
Center code: WUGSC			
Web site: http://genome.wustl.edu/gsc/index.shtml			
----- Project Information -----			
Center project name: H NH0585K18			
----- Summary Statistics -----			
Sequencing vector: M13; 71%			
Sequencing vector: plasmid; 5%			
Chemistry: Dye-terminator ET; 71% of reads			
Chemistry: Dye-terminator Big Dye; 5% of reads			
Assembly program: Phrap; version 0.990319			
Consensus quality: 143345 bases at least Q40			
Consensus quality: 146572 bases at least Q30			
Consensus quality: 148386 bases at least Q20			
Insert size: 164000; agarose-fp			
Insert size: 153491; sum-of-contigs			
Quality coverage: 4.65 in Q20 bases; agarose-fp			
Quality coverage: 5.04 in Q20 bases; sum-of-contigs			

```

ORIGIN
/note="assembly_name:Contig9"

Query Match      8.5%; Score 205; DB 2; Length 154791;
Best Local Similarity 100.0%; Pred. No. 2.5e-107;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 AGCCCTAGGTGATGGCATCGCTCTCTCACAGAAAAGTTCTTTTCCCATCTGAGAAGATTG 371
Db 73243 AGCCCTAGGTGATGGCATCGCTCTCTCACAGAAAAGTTCTTTTCCCATCTGAGAAGATTG 73184
QY 372 TCTTTAACTGGCAACAAACTCATAGAGTTGAGCTGGGCTCCAGAAATTCGGCAATACCTG 431
Db 73183 TCTTTAACTGGCAACAAACTCATAGAGTTGAGCTGGGCTCCAGAAATTCGGCAATACCTG 73124
QY 432 TTTTGGCAATGAGCACTGCGAGTGTTTAACTTACCTACACACACCTCTTTGCCAATTACATGCT 491
Db 73123 TTTTGGCAATGAGCACTGCGAGTGTTTAACTTACCTACACACACCTCTTTGCCAATTACATGCT 73064
QY 492 ATCACATGAACACTCCAAACATGT 516
Db 73063 ATCACATGAACACTCCAAACATGT 73039

RESULT 19
AX495102
LOCUS AX495102 213 bp DNA linear PAT 26-SEP-2002
DEFINITION Sequence 867 from Patent WO02059256.
ACCESSION AX495102
VERSION AX495102.1 GI:23340712
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1
AUTHORS Tuijinder,M., Teclerman,A., Amson,R. and Susini,L.
TITLE Sequences involved in phenomena of tumour suppression, tumour
reversion, apoptosis and/or virus resistance and their use as
medicines
JOURNAL Patent: WO 02059256-A 867 01-AUG-2002;
MOLECULAR ENGINEERING LAB (FR)
FEATURES
source
1..213
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match      5.7%; Score 137; DB 6; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.3e-67;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 GCTTCATGGAGAAACTCGTACAAGTCAGCAAGTGATGAAAGATGGTTCCAGCTTCAAA 977
Db 70 GCTTCATGGAGAAACTCGTACAAGTCAGCAAGTGATGAAAGATGGTTCCAGCTTCAAA 129
QY 978 GAGGTTCCTACTATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTTCGAA 1037
Db 130 GAGGTTCCTACTATCCATAGATCCTCTAATGTTCTTACACTTTCTCTGAAACGTTTTCGAA 189
QY 1038 TTTTACCGGTGGA AAAA 1054
Db 190 TTTTACCGGTGGA AAAA 206

RESULT 20
AX387329/c
LOCUS AX387329 292 bp DNA linear PAT 19-MAR-2002
DEFINITION Sequence 2257 from Patent WO0214500.
ACCESSION AX387329
VERSION AX387329.1 GI:19580459
KEYWORDS

```

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Escobedo, J., Garcia, P.D., Sudduth-Klinger, J., Reinhard, C.,
Randazzo, F., Lamson, G., Scott, E.M., Zhang, G., Kassam, A., Pot, D. and
Labat, I.
TITLE Human genes and gene expression products
JOURNAL Patent: WO 0214500-A 2257 21-FEB-2002;
CHIRON CORPORATION (US); Hyseq Inc. (US)
FEATURES
source Location/Qualifiers
1.292
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 4.1%; Score 100; DB 6; Length 292;
Best Local Similarity 99.3%; Pred. No. 6.2e-46;
Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 312 AGCCCTAGTGTGATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 371
Dd |||||
151 AGCCCTAGTGTGATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 92
QY 372 TCTTAAGTGGCAACAACACTCATAGTGTGAGCTGGCTCCAGAAATTTGGGCAATACCTG 431
Dd |||||
91 TCTTAAGTGGCAACAACACTCATAGTGTGAGCTGGCTCCAGAAATTTGGGCAATACCTG 32
QY 432 TTTTGGCAATGCGACTGCGAGTGTAAAC 462
Dd |||||
31 TTTTGGCAATGCGACTGCGAGTGTAAAC 1
RESULT 21
AC136746 195906 bp DNA linear HTG 10-MAR-2003
LOCUS Mus musculus clone RP23-151C5, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION pieces.
ACCESSION AC136746
VERSION HTG3 PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barn, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarata, J., Chang, V., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., DeAtrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N.,
Hagos, B., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J.,
Mathews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R.,
Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R.,
Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K.,
Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X.,
Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (07-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
TITLE
JOURNAL
AUTHORS

REFERENCE
AUTHORS 3 (bases 1 to 195906)
Birren, B., Nusbaum, C., Lander, E., Abouelheil, A., Allen, N.,
Anderson, S., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T.,
Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,
Collymore, A., Cook, A., Cooke, P., Corum, B., DeAtrellano, K.,
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B.,
Hall, J., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., MacLean, C.,
Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,
Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J.,
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C.,
Spencer, B., Stange-Thomann, N., Stojanovic, N., Topham, K., Wu, X.,
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wyman, D.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
COMMENT On Mar 10, 2003 this sequence version replaced gi:28882377.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L26341
Center clone name: 151C_5
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 194971 bases at least Q40
Consensus quality: 195140 bases at least Q30
Consensus quality: 195253 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 195506; sum-of-contigs
Quality coverage: 11.9 in Q20 bases; agarose-fp
Quality coverage: 11.4 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 104550: contig of 104650 bp in length
* 104651 104750: gap of 100 bp
* 104751 119592: contig of 14942 bp in length
* 119593 119792: gap of 100 bp
* 119793 138396: contig of 13204 bp in length
* 138397 139396: gap of 100 bp
* 139397 192208: contig of 53112 bp in length
* 192209 192308: gap of 100 bp
* 192309 195306: contig of 3598 bp in length.
* Location/Qualifiers
1.195906
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP23-151C5"
/clone.lib="RPCI-23 Female Mouse BAC"
1.104650
misc_feature

[illegible]

clone_end:T7
vector_side:right

ORIGIN

Query Match 1.4%; Score 34; DB 2; Length 195906;
Best Local Similarity 100.0%; Pred. No. 4,5e-07;
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 641 TCGGTTTGAACCAAGAGATGCCCATGAATT 674
|||||
Db 99247 TCGGTTTGAACCAAGAGATGCCCATGAATT 99214
|||||

RESULT 23
AC108995/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-298L20, *** SEQUENCING IN PROGRESS
***, 5 unordered clones.
AC108995 GI:25006701
VERSION HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 239456)
Muzny,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Chen,Y., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshewa,L., Loulseghe,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Mundaasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackelemech,O., Okwono,G., Olarnpusagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfamkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,M., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

Yu, F., Zhang, J., Zhou, X., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.
Direct Submission
Unpublished
2 (bases 1 to 239456)
Worley, K. C.
Direct Submission
Submitted (03-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 239456)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 15, 2002 this sequence version replaced gi:23195556.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GPNT
Center clone name: CH230-298L20
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 22813 bases at least Q40
Consensus quality: 22529 bases at least Q30
Consensus quality: 226929 bases at least Q20
Estimated insert size: 228958; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 216639: contig of 216639 bp in length
* 216640 216739: gap of unknown length
* 216740 224628: contig of 7889 bp in length
* 224629 224728: gap of unknown length
* 224729 234582: contig of 9854 bp in length
* 234583 234582: gap of unknown length
* 234583 235702: contig of 1020 bp in length
* 235703 235802: gap of unknown length
* 235803 239456: contig of 3654 bp in length.
----- Location/Qualifiers
1 .239456
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-298L20"
216740 .218521
/note="wgs_contig"

FEATURES
source
misc_feature

```

misc_feature 224729..226063
ORIGIN
/Note="wgs_contig"

Query Match 1.3% Score 32; DB 2; Length 239456;
Best Local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 412 CAGAAATTTGGCAATACCTCTTTTGGCAATGC 443
|||||
Db 150098 CAGAAATTTGGCAATACCTCTTTTGGCAATGC 150067

RESULT 24
AC111575/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-68E4, WORKING DRAFT SEQUENCE, 8
unordered pieces.
AC111575
VERSION
AC111575.4 GI:24818943
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 270456)
Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Derramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregregis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louised,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwakoelameh,O., Okwuonu,G., Olampunagoo,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabot,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Umani,K.,
Valas,R., Vera,V., Villasanad,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```

Yu, P., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstock, G. and Gibbs, R. A.

Unpublished
2 (bases 1 to 270456)
Worley, K. C.

Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 270456)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (09-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 9, 2002 this sequence version replaced gi:23267991.

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GMPC

Center clone name: CH230-68E4

Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 248347 bases at least Q40

Consensus quality: 250800 bases at least Q30

Consensus quality: 252288 bases at least Q20

Estimated insert size: 248995; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

	1	45421:	contig of 45421 bp in length
	45422	45521:	gap of unknown length
	45522	49934:	contig of 4413 bp in length
	49935	50034:	gap of unknown length
	50035	253199:	contig of 203165 bp in length
	253200	253299:	gap of unknown length
	253300	254387:	contig of 1088 bp in length
	254388	254487:	gap of unknown length
	254488	255765:	contig of 1278 bp in length
	255766	255865:	gap of unknown length
	255866	257214:	contig of 1349 bp in length
	257215	257314:	gap of unknown length
	257315	258888:	contig of 1574 bp in length
	258889	259888:	gap of unknown length
	258989	270456:	contig of 11468 bp in length.

Location/Qualifiers

1. 270456

FEATURES
source

RESULT 26

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Jones, K.A., Volkmut, W. and Walker, M.G.
TITLE Atherosclerosis-associated genes
JOURNAL Patent: WO 0104264-A 17 18-JAN-2001;
Incye Genomics, Inc. (US)
FEATURES
source
1..972
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 972;
Best Local Similarity 100.0%; Pred. No. 7.2e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTAGTGAGGGTTAATTTAGCTTGCA 2412
|||||
Db 936 CCCTTAGTGAGGGTTAATTTAGCTTGCA 965
|||||
RESULT 29
AX135572/c 1106 bp DNA linear PAT 29-MAY-2001
LOCUS
DEFINITION Sequence 69 from Patent WO0132888.
ACCESSION AX135572
VERSION AX135572.1 GI:14271842
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Yue, H., Hillman, J.L., Lal, P., Bandman, O., Patterson, C.,
Shih, L.L., Azimzai, Y., Lu, D.A. and Baughn, M.R.
TITLE Human transferase molecules
JOURNAL Patent: WO 0132888-A 69 10-MAY-2001;
Incye Genomics, Inc. (US)
FEATURES
source
1..1106
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incye ID No: 2448909CB1"
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1106;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTAGTGAGGGTTAATTTAGCTTGCA 2412
|||||
Db 38 CCCTTAGTGAGGGTTAATTTAGCTTGCA 9
|||||
RESULT 30
AX135583 1163 bp DNA linear PAT 29-MAY-2001
LOCUS
DEFINITION Sequence 80 from Patent WO0132888.
ACCESSION AX135583
VERSION AX135583.1 GI:14271853
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Tang, Y.T., Yue, H., Hillman, J.L., Lal, P., Bandman, O., Patterson, C.,
Shih, L.L., Azimzai, Y., Lu, D.A. and Baughn, M.R.
TITLE Human transferase molecules

JOURNAL Patent: WO 0132888-A 80 10-MAY-2001;
Incye Genomics, Inc. (US)
FEATURES
source
1..1163
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incye ID No: 4071304CB1"
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1163;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTAGTGAGGGTTAATTTAGCTTGCA 2412
|||||
Db 1127 CCCTTAGTGAGGGTTAATTTAGCTTGCA 1156
|||||
RESULT 31
AF121858 1487 bp mRNA linear PRI 14-SEP-2001
LOCUS
DEFINITION Homo sapiens sorting nexin 8 (SNX8) mRNA, complete cds.
ACCESSION AF121858
VERSION AF121858.1 GI:4689255
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1487)
AUTHORS Teasdale, R.D., Locci, D., Houghton, F., Karlsson, L. and Gleeson, P.A.
TITLE A large family of endosome-localized proteins related to sorting nexin 1
JOURNAL Biochem. J. 358 (Pt 1), 7-16 (2001)
MEDLINE 21378165
PUBMED 11485546
REFERENCE 2 (bases 1 to 1487)
AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.
TITLE Identification of eleven novel human sorting nexin molecules. A sub-group of the sorting nexin family is associated with the early endosomes
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1487)
AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-1999) The R.W. Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, CA 92121, USA
FEATURES
source
1..1487
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1487
/gene="SNX8"
11..1408
/gene="SNX8"
/codon_start=1
/product="sorting nexin 8"
/protein_id="AAD27831.1"
/db_xref="GI:4689256"
/translation="MTGRAMPPLPAAAVGAAAEADEADPPADSLPTPOAIFPOAI
VQVFPAPSRMQPQGNFLLSHTLQELLARDTVQVELIPEKGLFLKHVEYEVSSQRF
KSSVRYRNDVFWQEMLLHKFPYRMVVPALPKMKMGADREFIARRRALKRFVNLVA
RHPLFSDVLVLFSPGSDVONKLKESACVGVDEFNCKLATKADFLPADIQAF
AISELEIYNSEHKLDRARERTASRAIDNAADLLIFGKELSAIGSDTTPLPWAAL
NSSTWGLSKQALKGLSVEFALLADKAAQQQEENDVVEKLNLFDLLDLSQYKDLCERH
EKGVLHKHQALHKLKYSMLKRMQMSATQNPESVQEQLSEIVQEENAIQTMELENYF
SLYCLHQETQLIHVLPILTSLILRAFNVSQTQGHKEMSKVWMDLRPKSLCLFAGPHST
LTPCSPSPEDGLCPH"
ORIGIN
Query Match 1.2%; Score 30; DB 9; Length 1487;

Wed Aug 18 13:52:58 2004

Best Local Similarity 100.0%; Pred. No. 7.4e-05; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 |||||
 Db 1451 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 1480

RESULT 32
 AX061211
 LOCUS 1491 bp DNA linear PAT 22-JAN-2001
 DEFINITION Sequence 58 from Patent WO0078953.
 ACCESSION AX061211
 VERSION AX061211.1 GI:12406347
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Lal, P., Yang, J., Yue, H., Hillman, J.L., Tang, Y.T., Bandman, O.,
 Burford, N., Baughn, M.R., Azimzai, Y., Lu, D.A., Au-Young, J. and
 Patterson, C.
 TITLE Human transport proteins
 JOURNAL Patent: WO 0078953-A 58 28-DEC-2000;
 Incyte Genomics, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1491
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 1.2%; Score 30; DB 6; Length 1491;
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 |||||
 Db 1452 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 1481

RESULT 33
 AX073668
 LOCUS 1702 bp DNA linear PAT 06-FEB-2001
 DEFINITION Sequence 2 from Patent WO0104264.
 ACCESSION AX073668
 VERSION AX073668.1 GI:12710089
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Jones, K.A., Volkmut, W. and Walker, M.G.
 TITLE Atherosclerosis-associated genes
 JOURNAL Patent: WO 0104264-A 2 18-JAN-2001;
 Incyte Genomics, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1702
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 1.2%; Score 30; DB 6; Length 1702;
 Best Local Similarity 100.0%; Pred. No. 7.5e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 |||||
 Db 1665 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 1694

RESULT 34
 AX081430
 LOCUS 1950 bp DNA linear PAT 27-FEB-2001
 DEFINITION Sequence 14 from Patent WO0109178.
 ACCESSION AX081430
 VERSION AX081430.1 GI:13170253
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Yue, H., Bandman, O., Tang, Y.T., Baughn, M.R., Azimzai, Y. and Lu, D.A.
 TITLE Human chaperone proteins
 JOURNAL Patent: WO 0109178-A 14 08-FEB-2001;
 Incyte Genomics, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1950
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 1742550CB1"

ORIGIN
 Query Match 1.2%; Score 30; DB 6; Length 1950;
 Best Local Similarity 100.0%; Pred. No. 7.5e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 |||||
 Db 1915 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 1944

RESULT 35
 AX077793/c
 LOCUS 1954 bp DNA linear PAT 22-FEB-2001
 DEFINITION Sequence 24 from Patent WO0107628.
 ACCESSION AX077793
 VERSION AX077793.1 GI:1157660
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
 AUTHORS Tang, Y.T., Hillman, J.L., Bandman, O., Yue, H., Baughn, M.R., Lal, P.,
 Lu, D.A., Shah, P. and Azimzai, Y.
 TITLE Human synthetase
 JOURNAL Patent: WO 0107628-A 24 01-FEB-2001;
 Incyte Genomics, Inc. (US)
 FEATURES Location/Qualifiers
 source 1..1954
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 /note="Incyte ID No: 3231586CB1"

ORIGIN
 Query Match 1.2%; Score 30; DB 6; Length 1954;
 Best Local Similarity 100.0%; Pred. No. 7.5e-05;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
 |||||
 Db 48 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 19

RESULT 36
 BD260129/c
 LOCUS 1987 bp DNA linear PAT 17-JUL-2003
 DEFINITION Matrix-remodeling genes.
 ACCESSION BD260129

```

VERSION      BD260129.1  GI:33069899
KEYWORDS     JP 2002527054-A/20
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        Walker M.G., Volkmut, W. and Klingler, T.M.
JOURNAL      Matrix-remodeling genes
Patent: JP 2002527054-A 20 27-AUG-2002;
INCYTE PHARMACEUTICALS INC
COMMENT      OS Homo sapiens (human)
PN JP 2002527054-A/20
PD 27-AUG-2002
PF 06-OCT-1999 JP 2000575891
PR 09-OCT-1998 US 09/169289
P1 MICHAEL G WALKER, WAYNE VOLKUT, TOD M KLINGLER PC
C12N15/09, A61K38/00, A61K38/28, A61K48/46, A61K48/00, A61P1/04, PC
A61P3/10,
PC A61P9/10, A61P17/02, A61P19/02, A61P35/00, C07K14/47, C07K16/18, PC
C12N1/15,
PC C12N1/19, C12N1/21, C12N5/10, C12Q1/68, G01N33/15, G01N33/50, G01N33/53, G01N33/566, C12N15/00, C12N5/00, A61K37/02 CC
33,
PC G01N33/53//A61K31/711, A61K35/76, C12N15/00, C12N5/00, A61K37/02,
PC A61K37/26,
PC A61K37/54,
CC Incyte ID NO.: 3948614CB1
FH Key Location/Qualifiers
FT source 1..1987
FT /organism='Homo sapiens (human)'.
FEATURES
source 1..1987
Location/Qualifiers
/organism='Homo sapiens'
/mol_type='genomic DNA'
/db_xref='taxon:9606'
ORIGIN
Query Match 1.2%; Score 30; DB 6; Length 1987;
Best Local Similarity 100.0%; Pred. No. 7.6e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTTAGTGAGGGTTAATTTAGCTTGCA 2412
Db 39 CCCTTTAGTGAGGGTTAATTTAGCTTGCA 10
RESULT 37
BD224624/c
LOCUS      BD224624      3465 bp      DNA      linear      PAT 17-JUL-2003
DEFINITION Protein transport-associated molecules.
ACCESSION  BD224624
VERSION     BD224624.1  GI:33034394
KEYWORDS   JP 2002523089-A/4.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE        Tang, Y.T., Lal, P., Bandman, O., Yue, H., Corley, N.C., Guegl, K.J.,
JOURNAL      Gorgone, G.A., Baughn, M.R. and Patterson, C.
Patent: JP 2002523089-A 4 30-JUL-2002;
INCYTE PHARMACEUTICALS INC
COMMENT      OS Homo sapiens (human)
PN JP 2002523089-A/4
PD 30-JUL-2002
PF 26-AUG-1999 JP 2000567690
PR 27-AUG-1998 US 60/098206
P1 Y TOM TANG, PREETI LAL, OLGA BANDMAN, HENRY YUE, NEIL C CORLEY, PI
KARL J GUEGLER,
PI GINA A GORGONE, MARIAH R BAUGHN, CHANDRA PATTERSON PC
C12N15/09, A61K38/00, A61K39/395, A61K39/395, A61K39/395, A61K45/00, PC

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 20:20:52 ; Search time 1355 Seconds
(without alignments)
7562.106 Million cell updates/sec

Title: US-10-049-745-31

Perfect score: 2412

Sequence: 1 gggctgtgcggcgccgcg.....agggtaatttagcttgca 2412

Scoring table: OLIGO_NUC

Gapop_60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 30

Total number of hits satisfying chosen parameters: 69

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N Geneseq_29Jan04:*

1: geneseqn1980s:*

2: geneseqn1990s:*

3: geneseqn2000s:*

4: geneseqn2001as:*

5: geneseqn2001bs:*

6: geneseqn2002s:*

7: geneseqn2003as:*

8: geneseqn2003bs:*

9: geneseqn2003cs:*

10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	2412	100.0	2412	4	AAP81717	Aaf81717	Human pro
2	2321	96.2	4659	6	ABK86140	CDNA enco	
3	2287	94.8	2347	4	AHH19322	Human ubi	
4	2287	94.8	2347	7	ABX93880	Human CDN	
5	2265	93.9	3660	6	AHH17633	Human CDN	
6	2210	91.6	4483	6	ABZ11508	Human pol	
7	1475	61.2	1679	6	ABL39772	Human NS	
8	722	29.9	16682	4	AAK79798	Human imm	
9	722	29.9	16682	4	AAK70010	Human imm	
10	722	29.9	21436	4	AAK79799	Human imm	
11	722	29.9	21436	4	AAK70011	Human imm	
12	516	21.4	1122	4	AAS41077	CDNA enco	
13	424	17.6	577	4	AAS41613	CDNA enco	
14	411	17.0	411	4	AAI92171	Human pol	
15	357	14.8	722	6	ABK30453	Human G-p	
16	333	13.8	333	4	AAI22307	Probe #12	
17	333	13.8	333	4	ABA67385	Human foe	
18	333	13.8	333	4	ABA49470	Human bre	
19	333	13.8	333	4	AAK41552	Human bre	
20	333	13.8	333	5	AAI07994	Human bre	
21	333	13.8	333	6	ABK15563	Human gen	
22	314	13.0	498	4	AAI13071	Probe #30	
23	314	13.0	498	4	ABA54770	Human foe	

C 24	314	13.0	498	4	ABA44316	Aba44316	Human bre
C 25	314	13.0	498	4	AAK28495	Human bon	
C 26	314	13.0	498	5	AAI02976	Probe #29	
C 27	314	13.0	498	6	ABS03008	Human gen	
C 28	232	9.6	521	4	AAH05659	Human CDN	
C 29	205	8.5	457	6	ABK30444	Human G-p	
C 30	205	8.5	489	4	AAI15006	Human foe	
C 31	205	8.5	489	4	ABA56739	Human gen	
C 32	205	8.5	489	4	AAI36351	Probe #49	
C 33	205	8.5	489	4	ABA46202	Human foe	
C 34	205	8.5	489	4	AAK30399	Human bre	
C 35	205	8.5	489	4	ABS30038	Human bon	
C 36	205	8.5	489	5	AAI04774	Human liv	
C 37	205	8.5	489	6	ABS05002	Human gen	
C 38	198	8.2	198	4	AAI24193	Probe #14	
C 39	198	8.2	198	4	ABA69312	Human foe	
C 40	198	8.2	198	4	AAI49479	Human bre	
C 41	198	8.2	198	4	ABA51313	Human bon	
C 42	198	8.2	198	4	AAK43414	Human liv	
C 43	198	8.2	198	4	ABS43035	Human bon	
C 44	198	8.2	198	5	AAI09758	Human liv	
C 45	198	8.2	198	5	ABS17512	Probe #97	
C 46	137	5.7	213	6	ABQ95217	Human gen	
C 47	100	4.1	232	6	ABN62290	Tumour su	
C 48	85	3.5	657	4	AAK57307	Human can	
C 49	30	1.2	637	6	AAAD24377	Human imm	
C 50	30	1.2	733	4	AAAD08062	Human RNA	
C 51	30	1.2	823	4	AAAF59605	Human cel	
C 52	30	1.2	847	7	ACA04592	Human cel	
C 53	30	1.2	960	3	AAZ50898	CDNA enco	
C 54	30	1.2	972	4	AAAC5081	Human rec	
C 55	30	1.2	1106	5	AAH23827	Atheroscl	
C 56	30	1.2	1112	6	AAH42857	Human tra	
C 57	30	1.2	1163	5	AAH23838	Human DNA	
C 58	30	1.2	1273	8	ADA11069	Human tra	
C 59	30	1.2	1273	10	ADE77026	Human CDN	
C 60	30	1.2	1491	4	AAAF27715	Human tra	
C 61	30	1.2	1607	3	AAAF77351	Human ORF	
C 62	30	1.2	1702	4	AAAC85066	Human ORF	
C 63	30	1.2	1833	6	AAAD42873	Atheroscl	
C 64	30	1.2	1926	9	ADB47516	Human DNA	
C 65	30	1.2	1950	4	AAAF54986	Human CDN	
C 66	30	1.2	1954	4	AAAF60009	Nucleotid	
C 67	30	1.2	1987	3	AAZ52368	Human syn	
C 68	30	1.2	1987	6	AAZ52368	NSEQ gene	
C 69	30	1.2	3465	3	AAA08038	Human mat	

ALIGNMENTS

RESULT 1

AAAF81717

ID AAP81717 standard; CDNA; 2412 BP.

XX

AC AAP81717;

XX

DT 12-JUN-2001 (first entry)

XX

DE Human protease and protease inhibitor PPTM-4 encoding CDNA.

XX

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

KW

Human; protease; protease inhibitor; protease and protease inhibitor; PPTM; identification; diagnosis; anti-human immunodeficiency virus; HIV; antidiabetic; immunostimulant; immunomodulator; antiinflammatory; antithyroid; immunosuppressive; nephrotropic; antitumor; thyromimetic; cytostatic; antibacterial; fungicide; protozoacide; antiarteriosclerotic; antiatherosclerotic; antipsoriatic; virucide; hepatotropic; gene therapy; autoimmune disorder; inflammatory disease; AIDS; Chediak-Higashi syndrome; severe combined immunodeficiency disease; SCID; Crohn's disease; Addison's disease; diabetes mellitus; autoimmune thyroiditis; gout; Cushing's disease; Hashimoto's thyroiditis; Sjogren's syndrome; cancer; Grave's diseases; Hashimoto's thyroiditis; cell proliferative disorder; arteriosclerosis; Werner's syndrome;

1381 ATCAGTCAGCGGGTTGTCACCAACAAACAGCGCTGCGCGAGGCTTTATCGGACACACAGCTT 1440
1381 ATCAGTCAGCGGGTTGTCACCAACAAACAGCGCTGCGCGAGGCTTTATCGGACACACAGCTT 1440
1441 CCTCTCACATGATAAAGAAATCAACCTCTCACTTAAATGGGACTGGACCAATGAAAGACACG 1500
1441 CCTCTCACATGATAAAGAAATCAACCTCTCACTTAAATGGGACTGGACCAATGAAAGACACG 1500
1501 CGAAGCAGTTCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTCTT 1560
1501 CGAAGCAGTTCATGTCGAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTCTT 1560
1561 AATGCTTCAGTCTCTGTCACAAACTGTCAGTTAAATAGTTCCTCAGTGATCCAGAACAT 1620
1561 AATGCTTCAGTCTCTGTCACAAACTGTCAGTTAAATAGTTCCTCAGTGATCCAGAACAT 1620
1621 CCTAAGAAACAAAAAATTAACAATCAGTATTCACAAAGTTGCTGTTCCGCAAGTGTGAC 1680
1621 CCTAAGAAACAAAAAATTAACAATCAGTATTCACAAAGTTGCTGTTCCGCAAGTGTGAC 1680
1681 TCTCAACCTTAACCTTCATAGTATTTCTTTGGAGAAACCTTACCAAGCCGTTCCCTCTTCT 1740
1681 TCTCAACCTTAACCTTCATAGTATTTCTTTGGAGAAACCTTACCAAGCCGTTCCCTCTTCT 1740
1741 ACCATTACCAATTTCTGAGTACAGTCTACTCGAAGCATCTACGATGTCAGTTTCTAGT 1800
1741 ACCATTACCAATTTCTGAGTACAGTCTACTCGAAGCATCTACGATGTCAGTTTCTAGT 1800
1801 AAAAGTAAACAAACCGATCCCGCAGTGAATCTGCTCCAGCCGCTGATGAATGCAAA 1860
1801 AAAAGTAAACAAACCGATCCCGCAGTGAATCTGCTCCAGCCGCTGATGAATGCAAA 1860
1861 TCCAAGCTGAATCTCAGCTGCTGTCGCTATGCGCGAGTCTCTCAGGACTCTGAC 1920
1861 TCCAAGCTGAATCTCAGCTGCTGTCGCTATGCGCGAGTCTCTCAGGACTCTGAC 1920
1921 GAGGAGTCAAAAGGGCTGGCGAAGAGATGGATGTTGTTGATGTCAGTCCCACTCT 1980
1921 GAGGAGTCAAAAGGGCTGGCGAAGAGATGGATGTTGTTGATGTCAGTCCCACTCT 1980
1981 CCCGCCAAGATGCCAAGATGAGGAGGCACTCCGACAGCTTCAAGAACCCATGACC 2040
1981 CCCGCCAAGATGCCAAGATGAGGAGGCACTCCGACAGCTTCAAGAACCCATGACC 2040
2041 CTAACGGTGTCTAATAGTCAGACAGCGACAGTACCCGAAAGAAACCGCTTAGCGCT 2100
2041 CTAACGGTGTCTAATAGTCAGACAGCGACAGTACCCGAAAGAAACCGCTTAGCGCT 2100
2101 GATGTTGCCAGCTGCCAAGGCCAGCTGCCCTGCACTCAGAAATCCCTTTCGTAGGCA 2160
2101 GATGTTGCCAGCTGCCAAGGCCAGCTGCCCTGCACTCAGAAATCCCTTTCGTAGGCA 2160
2161 AACGCTTCTCTGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
2161 AACGCTTCTCTGAAAGTTGATGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2220
2221 TTAGAGACTTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAAATGATGCACT 2280
2221 TTAGAGACTTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAAATGATGCACT 2280
2281 GGAGCAGAGAGGGCCCTCCGAGACCGCGAGCTCAGCTGAGCGGAGCCCGCC 2340
2281 GGAGCAGAGAGGGCCCTCCGAGACCGCGAGCTCAGCTGAGCGGAGCCCGCC 2340
2341 GCCGAATCCCTGGAGGAGCCAGATGCGGCCGCAAGCTTATCCCTTTAGTGAGGTTAA 2400
2341 GCCGAATCCCTGGAGGAGCCAGATGCGGCCGCAAGCTTATCCCTTTAGTGAGGTTAA 2400
2401 TTTTACCTTGCA 2412
2401 TTTTACCTTGCA 2412

RESULT 2
ABK86140
ID ABK86140 standard; cDNA; 4659 BP.
XX AC ABK86140;
XX XX
DT 26-AUG-2002 (first entry)
XX cDNA encoding human protease PR7S7.
XX DE
XX KW Protease; human; gastrointestinal disorder; gastritis; atherosclerosis;
KW ulcerative colitis; Reye's syndrome; cardiovascular disorder; gene; ss;
KW hypertension; myocardial infarction; autoimmune disease; AIDS; PR7S;
KW inflammatory disorder; acquired immunodeficiency syndrome; asthma;
KW Grave's disease; cell proliferative disorder; hepatitis; psoriasis;
KW leukaemia; developmental disorder; Cushing's syndrome; impotence;
KW epithelial disorder; dermatitis; scabies; eczema; neurological disorder;
KW Parkinson's disease; dementia; Alzheimer's disease; infertility;
KW Huntington's disease; multiple sclerosis; reproductive disorder.
XX OS Homo sapiens.
XX XX
XX FH Location/Qualifiers
XX CDS 155..3910
XX FT /*tag= a
XX FT /product= "PR7S7 protein"
XX PN WO200238744-A2.
XX PD 16-MAY-2002.
XX XX
XX PF 18-OCT-2001; 2001WO-US051034.
XX PR 18-OCT-2000; 2000US-0241573P.
XX PR 25-OCT-2000; 2000US-0243643P.
XX PR 02-NOV-2000; 2000US-0245256P.
XX PR 13-NOV-2000; 2000US-0248395P.
XX PR 16-NOV-2000; 2000US-0249826P.
XX PR 20-NOV-2000; 2000US-0252303P.
XX PR 01-DEC-2000; 2000US-0250981P.
XX (INCY-) INCYTE GENOMICS INC.
XX PA Lee EA, Hafalia AJA, Yue H, Lal PG, Yao MG, Lu Y, Walia NK;
PI Warren BA, Lu DAM, Baughn MR, Deleane AM, Burford N, Borowsky ML;
PI Lee S, Xu Y, Griffin JA, Kallick DA, Gandhi AR, Arvizu C, Ison CH;
PI Tang YT, Azimzai Y, Elliott VS, Swarnakar A, Ramkumar J, Nguyen DB;
PI Tribouley CM, Lo TP, Au-Young J, Thangavelu K, Kearney L;
XX WPI; 2002-463471/49.
XX DR P-PSDB; AAU98889.
XX XX
XX PT New human proteases useful for diagnosing, preventing or treating
XX PT anorexia, myocardial infarction, Addison's disease, hepatitis, Cushing's
XX PT syndrome, eczema, Parkinson's disease, and impotence.
XX PS Claim 5; Page 161-162; 168pp; English.
XX XX
XX CC This invention relates to the DNA and protein sequences of novel isolated
XX CC human proteases (PR7S). the protein sequences of the invention are useful
XX CC for screening a compound for effectiveness as an agonist or antagonist of
XX CC its activity, the identified agonist and antagonist are useful for
XX CC treating a disease or condition associated with decreased or
XX CC overexpression of functional PR7S in a patient. The PR7S protein is also
XX CC useful as an immunogen for preparing polyclonal or monoclonal antibodies
XX CC by hybridoma technology. An antibody that binds the PR7S proteins is
XX CC useful for detection and purification of the proteins and can be used to
XX CC diagnose a condition or disease associated with expression of PR7S in a
XX CC subject or in a biological sample. The sequences of the invention are
XX CC useful for diagnosis, treatment and prevention of gastrointestinal
XX CC disorders such as gastritis, ulcerative colitis, Reye's syndrome, etc;
XX CC cardiovascular such as atherosclerosis, hypertension, myocardial


```
QY 622 CGCGTATAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAA 681
DB 610 CGCGGTATAGCTAGGACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAA 669
QY 682 TACACTGTTGATGCTATGTCAGAAAGATGCTTGAATGGCAGCAATAAAATAGACAGAC 741
DB 670 TACACTGTTGATGCTATGTCAGAAAGATGCTTGAATGGCAGCAATAAAATAGACAGAC 729
QY 742 ACCAGGCCACCACTCTTGTGTCAGATATTTGGAGGATACCTTAAGATCTAGAGTCAA 801
DB 730 ACCAGGCCACCACTCTTGTGTCAGATATTTGGAGGATACCTTAAGATCTAGAGTCAA 789
QY 802 TGTTTAAATTCGAAGGCGTTTCAGATACATTTTGGATCCATATCTTGATATAAATTCGAG 861
DB 790 TGTTTAAATTCGAAGGCGTTTCAGATACATTTTGGATCCATATCTTGATATAAATTCGAG 849
QY 862 ATAAAGGCTGCTCAGAGTGTCAACAAGGCATTTGAGCAGTGTGTAAGCCGGCAACAGCTT 921
DB 850 ATAAAGGCTGCTCAGAGTGTCAACAAGGCATTTGAGCAGTGTGTAAGCCGGCAACAGCTT 909
QY 922 GATGGAGAAATCTGTTAACAAGTCAGCAAGTGTAAGAGATGGTTCCAGCTTCAAAGAGG 981
DB 910 GATGGAGAAATCTGTTAACAAGTCAGCAAGTGTAAGAGATGGTTCCAGCTTCAAAGAGG 969
QY 982 TTCATATCCATAGATCCTCTAATGTTCTTACATTTCTCTGAAACGTTTTCGAAATTTT 1041
DB 970 TTCATATCCATAGATCCTCTAATGTTCTTACATTTCTCTGAAACGTTTTCGAAATTTT 1029
QY 1042 ACCGGTGGAAAAATTTGTAAGGATGTGAATAACCTCGAGTATCTTGATATTCGGCCATAT 1101
DB 1030 ACCGGTGGAAAAATTTGTAAGGATGTGAATAACCTCGAGTATCTTGATATTCGGCCATAT 1089
QY 1102 ATGTTCTCAACCCAAACGAGAGCCAAATTTGTAAGGATGTGAATAACCTCGAGTATCTTGATATTCGGCCATAT 1161
DB 1090 ATGTTCTCAACCCAAACGAGAGCCAAATTTGTAAGGATGTGAATAACCTCGAGTATCTTGATATTCGGCCATAT 1149
QY 1162 GGTTTTAAATTCGCAATGTCGCAATTTCTGCTACATATAAAGCTAGCAATGGCTCTCG 1221
DB 1150 GGTTTTAAATTCGCAATGTCGCAATTTCTGCTACATATAAAGCTAGCAATGGCTCTCG 1209
QY 1222 TATCAAAATGAATGACTCATTGTTATCTACCAAGTATATTAGATCGGTACTCAGCCAA 1281
DB 1210 TATCAAAATGAATGACTCATTGTTATCTACCAAGTATATTAGATCGGTACTCAGCCAA 1269
QY 1282 GCCTATGTCCTTTTATATCAGGTCCCATGATGTGAATAATGAGGTGAATCTACTCAT 1341
DB 1270 GCCTATGTCCTTTTATATCAGGTCCCATGATGTGAATAATGAGGTGAATCTACTCAT 1329
QY 1342 CCCACCATAGCCCGCCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGTTGTCA 1401
DB 1330 CCCACCATAGCCCGCCAGTCTCTCCCGCCCGCTCATCAGTCAGCGGTTGTCA 1389
QY 1402 AACAAACAGGCTGGCCAGGCTTTATCGGACCAACAGTTCCTCTCACATGATAAAGAT 1461
DB 1390 AACAAACAGGCTGGCCAGGCTTTATCGGACCAACAGTTCCTCTCACATGATAAAGAT 1449
QY 1462 CCACCTCACTTAATGAGCTGACCATTTGAAGACAGCCAGCAGTTCCTATGTCAGT 1521
DB 1450 CCACCTCACTTAATGAGCTGACCATTTGAAGACAGCCAGCAGTTCCTATGTCAGT 1509
QY 1522 CCTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTACGTTCTGTC 1581
DB 1510 CCTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTACGTTCTGTC 1569
QY 1582 AACTGGTCACTTAATGAGTCTTCAAGTATCCAGAACATCTTAAGAAACAAATAATCA 1641
DB 1570 AACTGGTCACTTAATGAGTCTTCAAGTATCCAGAACATCTTAAGAAACAAATAATCA 1629
QY 1642 ATCAGTATTCACAAAGTGTGCTTTCGCAAGTGTGCTACCTTCACTTCACTTATAGT 1701
DB 1630 ATCAGTATTCACAAAGTGTGCTTTCGCAAGTGTGCTACCTTCACTTCACTTATAGT 1689
QY 1702 AATTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTTACCATTTACCAATTTCTGCAGTA 1761
```

```
DB 1690 AATTTCTTGAGAAACCTTACCAAGCCGTTCCCTCTTCTACCATTTACCAATTTCTGCAGTA 1749
QY 1762 CAGTCTACCTCGAAGCATCTACGATGTCTAGTTTCTAGTAAAGTAAACAAACCGATCCCC 1821
DB 1750 CAGTCTACCTCGAAGCATCTACGATGTCTAGTTTCTAGTAAAGTAAACAAACCGATCCCC 1809
QY 1822 CGCAGTGAATCTCTGCTCCAGCCCGTATGAATGGCAATCCAAAGCTGAATCCAGCGTG 1881
DB 1810 CGCAGTGAATCTCTGCTCCAGCCCGTATGAATGGCAATCCAAAGCTGAATCCAGCGTG 1869
QY 1882 CTGCTGCTCTATGCGCCGAGTCTCTCTGAGGATCTCTGACAGGAGTCAAAAGGCGCTGGG 1941
DB 1870 CTGCTGCTCTATGCGCCGAGTCTCTCTGAGGATCTCTGACGAGGAGTCAAAAGGCGCTGGG 1929
QY 1942 AAGGAGATGGGATGTGATGATGTGAGTCTCCACTCTCCGCGCAAGATGCCGAAGAT 2001
DB 1930 AAGGAGATGGGATGTGATGATGTGAGTCTCCACTCTCCGCGCAAGATGCCGAAGAT 1989
QY 2002 GAGGAGGCCACTCCGCAAGGCTTCAAGAACCCATGACCCCTAAACGGTGTCTAATAGTGCA 2061
DB 1990 GAGGAGGCCACTCCGCAAGGCTTCAAGAACCCATGACCCCTAAACGGTGTCTAATAGTGCA 2049
QY 2062 GACAGCGACAGTGAACCGAAAGAAACGGCTAGCGCTGATGTGCTGAGTGCAGTGCAGAGG 2121
DB 2050 GACAGCGACAGTGAACCGAAAGAAACGGCTAGCGCTGATGTGCTGAGTGCAGTGCAGAGG 2109
QY 2122 CAGCTGCTCTGCTTCTGCTCTCTCCAGAAATCCCTTTGCTAAGCAACCGTCTTCTCTGGAAGTTG 2181
DB 2110 CAGCTGCTCTGCTTCTGCTCTCTCCAGAAATCCCTTTGCTAAGCAACCGTCTTCTCTGGAAGTTG 2169
QY 2182 ATGCTCTCTCTTCTGCTCTCTCCAGAAATCCCTTTGCTAAGCAACCGTCTTCTCTGGAAGTTG 2241
DB 2170 ATGCTCTCTTCTGCTCTCTCTCCAGAAATCCCTTTGCTAAGCAACCGTCTTCTCTGGAAGTTG 2229
QY 2242 AACAAATCTGAAAGCTCGACGGATGAATGAGTGCATCTGAGCAGAGAGAGGCGCTCTCC 2301
DB 2230 AACAAATCTGAAAGCTCGACGGATGAATGAGTGCATCTGAGCAGAGAGAGGCGCTCTCC 2289
QY 2302 GAGACCGCGACGCGGAGCTCAGCTCGGAGCCCGCCGCGCAATCCCTGGAGAGC 2359
DB 2290 GAGACCGCGACGCGGAGCTCAGCTCGGAGCCCGCCGCGCAATCCCTGGAGAGC 2347

RESULT 4
ABX93880
ID ABX93880 standard; cdna, 2347 BP.
XX
AC ABX93880;
XX
DT 10-JUN-2003 (first entry)
XX
DE Human cdna encoding ubiquitin protease.
XX
KW Human; ss; gene; ubiquitin protease; spleen disorder; Hodgkin's disease; lung disorder; adult respiratory distress syndrome; colon disorder; inflammatory bowel disease; liver disorder; jaundice; uterine disorder; endometriosis; brain disorder; Alzheimer's disease; T-cell disorder; acquired immunodeficiency syndrome; AIDS; skin disorder; urticaria; heart disorder; ischaemic heart disease; blood vessel disorder; atherosclerosis; red blood cell disorder; anaemia; thymus disorder; DiGeorge syndrome; B-cell disorder; leukaemia; kidney disorder; polycystic kidney disease; glomerulonephritis; breast disorder; mastitis; testicular disorder; sexually transmitted disease; thyroid disorder; hypothyroidism; pancreatic disorder; pancreatitis; intestinal disorder; whipple disease; tumour; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 61..2347
FT /*tag= a
FT /product= "Ubiquitin protease"
```

FT	/partial	/note= "No stop codon shown"	
FT XX	US2003037350-A1.		
FN XX	20-FEB-2003.		
PD XX			
XX XX	05-JUN-2002; 2002US-00163547.		
PF XX	29-SEP-1999; 99US-00407356.		
PR XX	05-NOV-1999; 99US-00435311.		
PR PR	01-FEB-2000; 2000US-00496005.		
PR PR	11-FEB-2000; 2000US-0182009P.		
PR PR	14-FEB-2000; 2000US-0182408P.		
PR PR	28-FEB-2000; 2000US-0185503P.		
PR PR	02-NOV-2000; 2000US-00704918.		
PR PR	12-FEB-2001; 2001US-00781598.		
PR PR	14-FEB-2001; 2001US-00782952.		
PR PR	28-FEB-2001; 2001US-00796100.		
XX XX	(MILL-) MILLENNIUM PHARM INC.		
PA XX	Glucksmann NA, Kapeller-Libermann R, Meyers RE, Rudolph-Owen LA;		
PI XX	WPI: 2003-342047/32.		
DR DR	P-PSDB; ABU08951.		
XX XX			
FT	New nucleic acid or polypeptide, useful for preparing a composition for		
FT	diagnosing or treating diseases e.g., tumor.		
XX XX			
PS	Claim 2; Fig 1; 281pp; English.		
XX XX			
CC	The invention relates to a new isolated nucleic acid molecule encoding		
CC	one of 6 polypeptides (ubiquitin protease, lipase, Dynamain, short chain		
CC	dehydrogenase, ADAM-TS (a disintegrin and metalloprotease domain protein		
CC	with thrombospondin (TS) domains) and gamma butyrobetaine-hydroxylase		
CC	(gamma BH). Also included are the polypeptide, host cells containing the		
CC	nucleic acids, an antibody that selectively binds to the polypeptide, a		
CC	method for producing the polypeptide, a method for detecting the presence		
CC	of the polypeptide or the nucleic acid in a sample, a method for		
CC	identifying a compound that binds to the polypeptide, a method for		
CC	modulating that modulates the activity of the polypeptide, a method for		
CC	or polypeptide is useful for preparing a composition for diagnosing or		
CC	treating diseases e.g. spleen disorders (e.g. splenomegaly and Hodgkin's		
CC	disease), lung disorders (e.g. adult respiratory distress syndrome, Hodgkin's		
CC	pulmonary oedema, chronic bronchitis and emphysema), colon disorders		
CC	(e.g. stenosis, colitis, inflammatory bowel disease and Crohn's disease),		
CC	liver disorders (e.g. jaundice, cirrhosis, hepatitis and alcoholic liver		
CC	disease), uterine and endometrial disorders (e.g. endometriosis and		
CC	menopausal changes), brain disorders (e.g. encephalitis, Alzheimer's		
CC	disease, Parkinson's disease, ataxia and multiple sclerosis), T-cell		
CC	disorders (e.g. acquired immunodeficiency syndrome, AIDS), skin disorders		
CC	(e.g. urticaria, dermatitis and lupus erythematosus), heart disorders		
CC	(e.g. ischaemic heart disease, myocardial infarction and cardiomyopathy),		
CC	blood vessel disorders (e.g. atherosclerosis, thrombophlebitis and		
CC	Raynaud disease), red blood cell disorders (e.g. anaemia), thymus		
CC	disorders (e.g. DiGeorge syndrome), B-cell disorders (e.g. leukaemia),		
CC	kidney disorders (e.g. polycystic kidney disease and glomerulonephritis),		
CC	breast disorders (e.g. mastitis), testicular disorders (e.g. sexually		
CC	transmitted diseases and cryptorchidism), thyroid disorders (e.g.		
CC	hypothyroidism), pancreatic disorders (e.g. pancreatitis), and intestinal		
CC	disorders (e.g. whipple disease), as well as tumours and cancers of the		
CC	above listed organs/cells. Many more diseases and disorders are listed in		
CC	the specification. The present sequence encodes human ubiquitin protease		
XX			
SQ	Sequence 2347 BP; 647 A; 606 C; 528 G; 566 T; 0 U; 0 Other;		
	Query Match 94.8%; Score 2287; DB 7; Length 2347;		
	Best Local Similarity 100.0%; Pred. No. 0;		
	Matches 2337; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	22 GCGGCGCGCGGGGATGGAGCGAGCGCGCGCGGGTCAGAGTTGACAAATGACCATTA 81		

Db 1090 ATGCTCAACCCACGAGAGCAATGTCTACGTCTTGTATCCAGTGTGGTCCACACT 1149
Qy 1162 GGTTTTAATTGCCATGCTGGCCATTACTTCTGTCTACATAAAAGCTAGCAATGGCTCTGG 1221
Db 1150 GGTTTTAATTGCCATGCTGGCCATTACTTCTGTCTACATAAAAGCTAGCAATGGCTCTGG 1209
Qy 1222 TATCAATGAATGACATCCATTTGATCTACAGTGAATATAGATCGGTACTCAGCCACAA 1281
Db 1210 TATCAATGAATGACATCCATTTGATCTACAGTGAATATAGATCGGTACTCAGCCACAA 1269
Qy 1282 GCCTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAAATGAGGTGAATCTACTCAT 1341
Db 1270 GCCTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAAATGAGGTGAATCTACTCAT 1329
Qy 1342 CCACCCATAGCCCGCCAGTCTCTCCCGCCCGTCTATCATGTCAGCGGGTGTGCACC 1401
Db 1330 CCACCCATAGCCCGCCAGTCTCTCCCGCCCGTCTATCATGTCAGCGGGTGTGCACC 1389
Qy 1402 AACAAACAGGCTGGCCAGGCTTTATCGGACACACAGCTTCCCTCTCATATGATAAAGAA 1461
Db 1390 AACAAACAGGCTGGCCAGGCTTTATCGGACACACAGCTTCCCTCTCATATGATAAAGAA 1449
Qy 1462 CCACCTCACCTTAATGAGTGGACCAATGAAGACACGCAAGCAAGTTCATGTTCAGT 1521
Db 1450 CCACCTCACCTTAATGAGTGGACCAATGAAGACACGCAAGCAAGTTCATGTTCAGT 1509
Qy 1522 CTTAACCGGAATTCAGTGTCAACAGGGCTAGTCTGTTAATGTTTCTGCTTCTGTTCCAA 1581
Db 1510 CTTAACCGGAATTCAGTGTCAACAGGGCTAGTCTGTTAATGTTTCTGCTTCTGTTCCAA 1569
Qy 1582 AACTGTCAGTAAATAGTCTCAGTATCCAGAACATCTTAAGAAACAAAAATTTACA 1641
Db 1570 AACTGTCAGTAAATAGTCTCAGTATCCAGAACATCTTAAGAAACAAAAATTTACA 1629
Qy 1642 ATCAGTATTCACAAACAGTTCCTGTTGCGCAGTGTCTAGTCTCAACCTTAACCTTCATAGT 1701
Db 1630 ATCAGTATTCACAAACAGTTCCTGTTGCGCAGTGTCTAGTCTCAACCTTAACCTTCATAGT 1689
Qy 1702 AATCTTTGGAGAACCTTACAGTGTCTAGTGTCTAGTAAAGTAAACAAACCGATCCCC 1821
Db 1750 CAGTCTACCTCGAAGCAGTATCAGATGTCAGTGTCTAGTAAAGTAAACAAACCGATCCCC 1809
Qy 1822 CGCAGTGAATCTGCTGCCAGCCGCTGATGAATGCAAAATCCAAAGTGAATCCAGCGGT 1881
Db 1810 CGCAGTGAATCTGCTGCCAGCCGCTGATGAATGCAAAATCCAAAGTGAATCCAGCGGT 1869
Qy 1882 CTGCTGCCCTATGGCCCGAGTCTCTGAGGACTCTGAGGAGTCAAGGGGCTGGC 1941
Db 1870 CTGCTGCCCTATGGCCCGAGTCTCTGAGGACTCTGAGGAGTCAAGGGGCTGGC 1929
Qy 1942 AAGAGAGTGGATGGTACGATGTGAGTCCCACTCTCCCGCCAGAGTCCCGAAGAT 2001
Db 1930 AAGAGAGTGGATGGTACGATGTGAGTCCCACTCTCCCGCCAGAGTCCCGAAGAT 1989
Qy 2002 GAGGAGGCCATCTCCGACAGAGTTCAGAAACCATATGACCCCTAAACGGTGTAAATAGTCA 2061
Db 1990 GAGGAGGCCATCTCCGACAGAGTTCAGAAACCATATGACCCCTAAACGGTGTAAATAGTCA 2049
Qy 2062 GACAGGACAGTGGCCGAAAGAAACCGCTAGCGCTGATGGTCCAGCTGCCAAGGC 2121
Db 2050 GACAGGACAGTGGCCGAAAGAAACCGCTAGCGCTGATGGTCCAGCTGCCAAGGC 2109
Qy 2122 CAGCTGCCCTGCATCTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCTCTGGAAGTTG 2181
Db 2110 CAGCTGCCCTGCATCTCAGAAATCCCTTTGCTAAGGCAACCGTCTTCTCTGGAAGTTG 2169
Qy 2182 ATGCTGCTCTCTTGTCTCTCCAGAGCAAAATCTTAGACACCTTCAGGCTTAGC 2241
Db 2170 ATGCTGCTCTCTTGTCTCTCTCCAGAGCAAAATCTTAGACACCTTCAGGCTTAGC 2229

Qy 2242 AACAAACTGAAGGCTCGACGGATGAATCAGTGCACCTGGACGAGAGGGGCCCTCCC 2301
Db 2230 AACAAACTGAAGGCTCGACGGATGAATCAGTGCACCTGGACGAGAGGGGCCCTCCC 2289
Qy 2302 GAGGACCGGACGCCGAGGCTCAGCTGCGAGCCCCCGCCGCAATCCCTGGAGGAGC 2359
Db 2290 GAGGACCGGACGCCGAGGCTCAGCTGCGAGCCCCCGCCGCAATCCCTGGAGGAGC 2347

RESULT 5
AAH17633

ID AAH17633 standard; cDNA; 3660 BP.

XX AAH17633;

XX 26-JUN-2001 (first entry)

XX Human cDNA sequence SEQ ID NO:17168.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-0030253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

XX Claim 8; SEQ ID NO 17168; 2537pp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dr primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX

QY	2106	TGCAGCTGCCAAGGCAGCCTCGCTTGCCTTGTCTCTCCAGAGAGA	2225	CGG	2160	TGCAGCTGCCAAGGCAGCCTCGCTTGTCTCTCCAGAGAGA	2225	CGG	2160
DB	2101	TGCAGCTGCCAAGGCAGCCTCGCTTGTCTCTCCAGAGAGA	2225	CGG	2160	TGCAGCTGCCAAGGCAGCCTCGCTTGTCTCTCCAGAGAGA	2225	CGG	2160
QY	2166	TCTTCTCGGAAAGTTGATGCTGCTTGTCTCTCCAGAGAGA	2225	CGG	2160	TCTTCTCGGAAAGTTGATGCTGCTTGTCTCTCCAGAGAGA	2225	CGG	2160
DB	2161	TCTTCTCGGAAAGTTGATGCTGCTTGTCTCTCCAGAGAGA	2225	CGG	2160	TCTTCTCGGAAAGTTGATGCTGCTTGTCTCTCCAGAGAGA	2225	CGG	2160
QY	2226	GACCTTCAGCTTAGCAACAACATGAAGGCTCGCGGATGAATGATGCTGAGC	2285	GAGC	2280	GACCTTCAGCTTAGCAACAACATGAAGGCTCGCGGATGAATGATGCTGAGC	2285	GAGC	2280
DB	2221	GACCTTCAGCTTAGCAACAACATGAAGGCTCGCGGATGAATGATGCTGAGC	2285	GAGC	2280	GACCTTCAGCTTAGCAACAACATGAAGGCTCGCGGATGAATGATGCTGAGC	2285	GAGC	2280
QY	2286	AGAGAGGGGCGCTCCCGAGGACCGGAGCGCGGAGCTCAGCCTGGCAGCCCGCGCGA	2345	CGG	2340	AGAGAGGGGCGCTCCCGAGGACCGGAGCGCGGAGCTCAGCCTGGCAGCCCGCGCGA	2345	CGG	2340
DB	2281	AGAGAGGGGCGCTCCCGAGGACCGGAGCGCGGAGCTCAGCCTGGCAGCCCGCGCGA	2345	CGG	2340	AGAGAGGGGCGCTCCCGAGGACCGGAGCGCGGAGCTCAGCCTGGCAGCCCGCGCGA	2345	CGG	2340
QY	2346	ATCCCTGGAGGAGCCAGATCGGCGCG	2372			ATCCCTGGAGGAGCCAGATCGGCGCG	2372		
DB	2341	ATCCCTGGAGGAGCCAGATCGGCGCG	2372			ATCCCTGGAGGAGCCAGATCGGCGCG	2372		
RESULT 6									
ABZ11508	ABZ11508 standard; cDNA; 4483 BP.								
ID									
XX									
AC	ABZ11508;								
XX									
DT	20-JAN-2003 (first entry)								
XX									
DE	Human polynucleotide SEQ ID NO 390.								
XX									
XX	Human; genome mapping; gene therapy; food supplement; virus; fungus;								
KW	cell-proliferative disorder; neurodegenerative disease; bacterial;								
KW	Parkinson's disease; Alzheimer's disease; autoimmune disease;								
KW	multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;								
KW	arthritis; cytostatic; immunomodulator; neoplastic; dermatological;								
KW	antiparkinsonian; antidiabetic; immunosuppressive; dermatological;								
KW	haemostatic; vulnery; fungicide; antibacterial; virucide; protozoacide;								
KW	antiarthritic; gene; ss.								
OS	Homo sapiens.								
XX									
XX	WO200270539-A2.								
PN									
XX									
PD	12-SEP-2002.								
XX									
PF	05-MAR-2002; 2002WO-US0005095.								
XX									
PR	05-MAR-2001; 2001US-00799451.								
XX									

CC	imaging, screening and diagnostic assays and for treating cell-
CC	proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC	or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC	diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
CC	platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC	or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC	arthritis, etc. Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX	
SQ	Sequence 4483 BP; 1127 A; 1270 C; 1169 G; 917 T; 0 U; 0 Other;
	Query Match 91.6%; Score 2210; DB 6; Length 4483;
	Best Local Similarity 99.9%; Pred No. 0;
	Matches 2310; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY	61 CAGAGTTGAACAATCACCATAGTTCACAAAGCTTCGAAATCTTCAGACCCATCACGCTAT 120
DB	179 CAGAGTTGAACAATCACCATAGTTCACAAAGCTTCGAAATCTTCAGACCCATCACGCTAT 238
QY	121 CAGAATCAGCGTGAGCTCCGAGGCAGTCTCACTGGAGACATGATGACAGTTCGTGCC 180
DB	239 CAGAATCAGCGTGAGCTCCGAGGCAGTCTCACTGGAGACATGATGACAGTTCGTGCC 298
QY	181 AGCTGGGGTGCTGTGTCCTCATTTGAATGATGTCAAATCA CACACTTTCTTTAGAACCA 240
DB	299 AGCTGGGGTGCTGTGTCCTCATTTGAATGATGTCAAATCA CACACTTTCTTTAGAACCA 358
QY	241 GTACCTGTGTCTGTAGTTATTTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA 300
DB	359 GTACCTGTGTCTGTAGTTATTTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA 418
QY	301 CAAAAGGATCAAGCCCTAGTGTATGGCATCGCTCCTCCACAGAAAGTCTCTTTTCCCATCT 360
DB	419 CAAAAGGATCAAGCCCTAGTGTATGGCATCGCTCCTCCACAGAAAGTCTCTTTTCCCATCT 478
QY	361 GAGAAGATTGTCTTAAGTGGCAACAAATCATAGAGTTGGAGCTGGGCTCCAGAAATTG 420
DB	479 GAGAAGATTGTCTTAAGTGGCAACAAATCATAGAGTTGGAGCTGGGCTCCAGAAATTG 538
QY	421 GGCAATACCTGTTTTGCCAATGACGACCTGCAGTGTTTAACTTACACACACCTCTTTGCC 480
DB	539 GGCAATACCTGTTTTGCCAATGACGACCTGCAGTGTTTAACTTACACACACCTCTTTGCC 598
QY	481 AAATTACATGCTATCACAATGAACACTCCAAACAAATGATCATGACAGAGCTTTGTATGATG 540
DB	599 AAATTACATGCTATCACAATGAACACTCCAAACAAATGATCATGACAGAGCTTTGTATGATG 658
QY	541 TGTTACAATGCAAGCANATTTACCCAGGCACTCAGTAATCTCTGGGACGCTTATTAACACCA 600
DB	659 TGTTACAATGCAAGCANATTTACCCAGGCACTCAGTAATCTCTGGGACGCTTATTAACACCA 718
QY	601 ATGTTTCTCATCAATGAGATGCGCGGTATAGCTAGGCACCTCCGTTTGTGAAACCAAGAA 660
DB	719 ATGTTTCTCATCAATGAGATGCGCGGTATAGCTAGGCACCTCCGTTTGTGAAACCAAGAA 778
QY	661 GATGCCCATGAAATTCCTTCAATACATCTGTTGATGCTATGCAGAAAGCATCTTGAATGCC 720
DB	779 GATGCCCATGAAATTCCTTCAATACATCTGTTGATGCTATGCAGAAAGCATCTTGAATGCC 838
QY	721 AGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTGTGTGCAGATATTTGGAGGA 780
DB	839 AGCAATAAATTAGACAGACACACCCAGGCCACCACTCTTGTGTGCAGATATTTGGAGGA 898
QY	781 TACCTAAGATCTTAGAGTCAAAATGTTTAAATGCAAGGCGGTTTCAGATACCTTTTGATCCA 840
DB	899 TACCTAAGATCTTAGAGTCAAAATGTTTAAATGCAAGGCGGTTTCAGATACCTTTTGATCCA 958
QY	841 TATCTTGATATAACATTTGGAGATAAAGCGTCTCAGAGTGTCAACAGGCAATTTGGAGCAG 900
DB	959 TATCTTGATATAACATTTGGAGATAAAGCGTCTCAGAGTGTCAACAGGCAATTTGGAGCAG 1018
QY	901 TTGTGTAGCGCGAACAGCTTTGATGGAGAAAACTCTGTACAAGTGCAGCAAGTGTAAAAAG 960

Db 1019 TTGTGAAGCCGGAACAGCTTGTATGAGAAAACCTGTCAGATGTCAGCAAGTGTAAAG 1078
QY 961 ATGGTTCCAGCTTCAAGAGAGTTCACTATCCATAGATCCCTCTAATGTTCTTACACTTCT 1020
Db 1079 ATGGTTCCAGCTTCAAGAGAGTTCACTATCCATAGATCCCTCTAATGTTCTTACACTTCT 1138
QY 1021 CTGAACGCTTTTCAAAATTTTACGGTGGAAAATTCCTAAGGATGTGAATACCCCTGAG 1080
Db 1139 CTGAACGCTTTTCAAAATTTTACGGTGGAAAATTCCTAAGGATGTGAATACCCCTGAG 1198
QY 1081 TATCTTGATATTCGGCCATATATGTCTCAACCCAAACGAGAGCAATGTCTACGCTTTC 1140
Db 1199 TATCTTGATATTCGGCCATATATGTCTCAACCCAAACGAGAGCAATGTCTACGCTTTC 1258
QY 1141 TATGAGTCTGCTGTCACACTGTTTAAATGCGATGCTGGCCATTAATGCTGCTCAAT 1200
Db 1259 TATGAGTCTGCTGTCACACTGTTTAAATGCGATGCTGGCCATTAATGCTGCTCAAT 1318
QY 1201 AAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCATTCATCTACCACTGATATT 1260
Db 1319 AAAGCTAGCAATGGCTCTGGTATCAAAATGAATGACTCCATTCATCTACCACTGATATT 1378
QY 1261 AGATCGGTACTCAGCCAAACAGCCTATGTCTCTTTTATATCAGGTCCCATGATGTGAAA 1320
Db 1379 AGATCGGTACTCAGCCAAACAGCCTATGTCTCTTTTATATCAGGTCCCATGATGTGAAA 1438
QY 1321 AATGGAGGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1380
Db 1439 AATGGAGGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1498
QY 1381 ATCAGTCAAGCGGTGTCACCAACAAACAGGCTGCGCCAGGCTTTATCGACACACAGTT 1440
Db 1499 ATCAGTCAAGCGGTGTCACCAACAAACAGGCTGCGCCAGGCTTTATCGACACACAGTT 1558
QY 1441 CCTCTCATGATTAAGAAATCACTCTAAATGGAGCTGGACCAATGAAGACACG 1500
Db 1559 CCTCTCATGATTAAGAAATCACTCTAAATGGAGCTGGACCAATGAAGACACG 1618
QY 1501 CCAACAGTTCATGTCGAGTCTTAACGGAATTCAGTGTCAACAGGCTAGTCTGTT 1560
Db 1619 CCAACAGTTCATGTCGAGTCTTAACGGAATTCAGTGTCAACAGGCTAGTCTGTT 1678
QY 1561 AATGCTTCAGTCTTCCTCAAACTGCTCAGTTAATAGTCTCAGTGTATCCAGAACAT 1620
Db 1679 AATGCTTCAGTCTTCCTCAAACTGCTCAGTTAATAGTCTCAGTGTATCCAGAACAT 1738
QY 1621 CTTAAGAAACAAAAATTAACAATCAGTATTCACAAAGTTGCTGTTGCGCAGTGTGAC 1680
Db 1739 CTTAAGAAACAAAAATTAACAATCAGTATTCACAAAGTTGCTGTTGCGCAGTGTGAC 1798
QY 1681 TCTCAACCTTAACCTTATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCT 1740
Db 1799 TCTCAACCTTAACCTTATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCT 1858
QY 1741 ACCATTACCAATTCCTGAGTACAGTCTACCTCGAACCGCATCTACGATGTCAGTTCTAGT 1800
Db 1859 ACCATTACCAATTCCTGAGTACAGTCTACCTCGAACCGCATCTACGATGTCAGTTCTAGT 1918
QY 1801 AAAGTAAACAAACCGATCCCGCGAGTGAATCCTGCTCCAGCCCGTGAATGAGTGAACAA 1860
Db 1919 AAAGTAAACAAACCGATCCCGCGAGTGAATCCTGCTCCAGCCCGTGAATGAGTGAACAA 1978
QY 1861 TCCAGAGTGAATCCAGCGTGTGCTGTCCTATGCGCGGAGTCTCTGAGGACTCTGAC 1920
Db 1979 TCCAGAGTGAATCCAGCGTGTGCTGTCCTATGCGCGGAGTCTCTGAGGACTCTGAC 2038
QY 1921 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGATTGTACGATGTGAGTCCCACTCT 1980
Db 2039 GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGATTGTACGATGTGAGTCCCACTCT 2098
QY 1981 CCGGCCAAGATGCCGAATGAGGAGGCCAATCCGACAGGCTTCAAGAACCCATGACC 2040

Db 2099 CCGGCCAAGATGCCGAATGAGGAGGCCACTCCGACAGCTTCAAGAACCCATGACC 2158
QY 2041 CTAACCGTGTCTAATAGTGCAGACAGCGACAGTACCCGAAAGAAAAACGGCTAGCGCCT 2100
Db 2159 CTAACCGTGTCTAATAGTGCAGACAGCGACAGTACCCGAAAGAAAAACGGCTAGCGCCT 2218
QY 2101 GATGGTCCAGCTGCCAAGCCAGCCCTGCCCTGCATCTAGAAAAATCCCTTTGCTAAGGCA 2160
Db 2219 GATGGTCCAGCTGCCAAGCCAGCCCTGCCCTGCATCTAGAAAAATCCCTTTGCTAAGGCA 2278
QY 2161 AACGGTCTTCTCTGAAAGTTGATGCTGCTCCTTTGCTGCTCTCCCAAGAACAAATC 2220
Db 2279 AACGGTCTTCTCTGAAAGTTGATGCTGCTCCTTTGCTGCTCTCCCAAGAACAAATC 2338
QY 2221 TTAGAACCTTTCAGGCTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGCACCT 2280
Db 2339 TTAGAACCTTTCAGGCTTAGCAACAAACTGAAAGGCTCGACGGATGAAATGAGTGCACCT 2398
QY 2281 GGAGCAGAGAGGGGCTCCCGAGGACCGGAGCGGAGCTCAGCTGCGACCGCGCC 2340
Db 2399 GGAGCAGAGAGGGGCTCCCGAGGACCGGAGCGGAGCTCAGCTGCGACCGCGCC 2458
QY 2341 GCCGAATCCTTGAGGAGCCAGATGCGCGCG 2372
Db 2459 GCCGAATCCTTGAGGAGCCAGATGCGCGCG 2490
RESULT 7
ABL39772
ID ABL39772 standard; cDNA; 1679 BP.
XX ABL39772;
XX
DT 10-MAY-2002 (first entry)
XX
DE Human NS cDNA sequence SEQ ID NO:82.
XX
KW Human; cytostatic; osteopathic; gynaecological; neuroprotective;
KW antirheumatic; antiarthritic; antipsoriatic; ophthalmological; anti-HIV;
KW vasotropic; antiarteriosclerotic; antiinflammatory; dermatological;
KW anorectic; muscular; antiinfertility; cardiovascular; anticoagulant;
KW antifibrinolytic; hypotension; antiasthmatic; immunomodulator; cardiac;
KW anticonvulsant; antidiabetic; tranquiliser; antidepressant; neuroleptic;
KW gastrointestinal; virucide; antiulcer; antidiabetic; cerebroprotective; nootropic;
KW contraceptive; vaccine; gene therapy; cancer; osteoporosis; dystonia;
KW endometriosis; degenerative disease; multiple sclerosis; psoriasis;
KW rheumatoid arthritis; cataract; restenosis; atherosclerosis; glaucoma;
KW infertility; skin disorder; obesity; muscular dystrophy; AIDS;
KW inflammation; cardiovascular disease; coagulation disease; hypertension;
KW ischaemia; asthma; immune disease; epilepsy; angina; neurodegeneration;
KW diabetes; anxiety; depression; schizophrenia; viral disease; stroke;
KW gastric ulcer; Alzheimer's disease; gene; ss.
XX Homo sapiens.
XX
XX WO200206315-A2.
XX
XX 24-JAN-2002.
XX
XX 17-JUL-2001; 2001WO-IL000653.
XX
XX 18-JUL-2000; 2000IL-00137345.
XX
XX 15-DEC-2000; 2000IL-00140354.
XX
XX (COMP-) COMPUGEN LTD.
XX
XX Mintz L, Freilich S, Bernstein J;
XX
XX WPI; 2002-155037/20.
XX
XX P-PSDB; ABB06118.
XX
XX One hundred and twenty eight novel nucleic acid sequences, useful for
XX treating and diagnosing e.g. cancer, asthma and Alzheimer's.

RESULT 8
AAK79798
ID AAK

XX
AC AAK79798;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34610.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
FN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 12-SEP-2000; 2000US-0232081P.
PR 14-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250160P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.

PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 05-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 06-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 21-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234224P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234999P.
PR 27-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 29-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 20-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.
Disclosure; SEQ ID NO 24822; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patients own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting the
nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/haematopoietic-related diseases, especially
cancers and cancer metastases of haematopoietic-derived cells. AAK64703
to AAK87694 represent human immune/haematopoietic antigen genomic
sequences from the present invention. AAK54942 to AAK54950 and AAM82169
represent sequences used in the exemplification of the present invention

XX SQ Sequence 16682 BP; 4217 A; 3754 C; 4167 G; 4538 T; 0 U; 6 Other;

Query Match 29.9%; Score 722; DB 4; Length 16682;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1457 AGATCCACCTCACTTAATGGGACTGACCATTTGAAGACACCCAGCAGTTCCATGT 1516

|||||

Db 9451 AGAATCCACCTCACITTAATGGAGCTGGACCATTGAAGACACGCCAAGCAGTTCCATCT 9510
QY 1517 CGAGTCCTTAACGGGAATCCAGTGTCAACAGGCTAGTCTGTAAATGCTTACGTTCTG 1576
Db 9511 CGAGTCCTTAACGGGAATCCAGTGTCAACAGGCTAGTCTGTAAATGCTTACGTTCTG 9570
QY 1577 TCCAAACTGGTCAGTTAATAGTCTCTCAGTATCCAGACATCTTAAGAACCAAAAA 1636
Db 9571 TCCAAACTGGTCAGTTAATAGTCTCTCAGTATCCAGACATCTTAAGAACCAAAAA 9630
QY 1637 TTACAATCAGTATTCACAACAGTTGCTCTTCGCGAGTGTCACTCAACCTTAACCTTC 1696
Db 9631 TTACAATCAGTATTCACAACAGTTGCTCTTCGCGAGTGTCACTCAACCTTAACCTTC 9690
QY 1697 ATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCCTTTCTACCATTAACAAATCTG 1756
Db 9691 ATAGTAATCTTTGGAGAACCTTACCAAGCCGTTCCCTTTCTACCATTAACAAATCTG 9750
QY 1757 CAGTACAGTCTACTCGAAGCGATCTACGATGTCAGTCTTCTAGTAAAGTAAACCA 1816
Db 9751 CAGTACAGTCTACTCGAAGCGATCTACGATGTCAGTCTTCTAGTAAAGTAAACCA 9810
QY 1817 TCCCCCGCAGTGAATCTGCTCCAGCCGCTGATGAATGGCAATCCAGCTGAACCTCA 1876
Db 9811 TCCCCCGCAGTGAATCTGCTCCAGCCGCTGATGAATGGCAATCCAGCTGAACCTCA 9870
QY 1877 GCGTGTGTCCTTATGGCCGAGTCTCTGAGGACTCTGACGAGGATCAAGGGGC 1936
Db 9871 GCGTGTGTCCTTATGGCCGAGTCTCTGAGGACTCTGACGAGGATCAAGGGGC 9930
QY 1937 TGGCAGAGGATGGATGGTACGATTTGTAGCTCCACTCCCGCCAGATCCG 1996
Db 9931 TGGCAGAGGATGGATGGTACGATTTGTAGCTCCACTCCCGCCAGATCCG 9990
QY 1997 AAGATGAGGAGGACCTCCGACAGCTTCAAGAACCCATGACCTTAACCGTCTAATA 2056
Db 9991 AAGATGAGGAGGACCTCCGACAGCTTCAAGAACCCATGACCTTAACCGTCTAATA 10050
QY 2057 GTGCAGACAGCAGTACCCGAAAGAACCGGCTAGCGCTGATGGTCCAGCTGCC 2116
Db 10051 GTGCAGACAGCAGTACCCGAAAGAACCGGCTAGCGCTGATGGTCCAGCTGCC 10110
QY 2117 AAGCCAGCTGCTCCACTCAGAAATCCCTTTGTAGGCAAAACGGTCTTCTCGAA 2176
Db 10111 AAGCCAGCTGCTCCACTCAGAAATCCCTTTGTAGGCAAAACGGTCTTCTCGAA 10170
QY 2177 AG 2178
Db 10171 AG 10172

RESULT 10
AAK79799
ID AAK79799 standard; DNA; 21436 BP.
XX
AC AAK79799;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:34611.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX
KW cytosolic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US001354.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR

PR 04-FEB-2000; 2000US-0180528P.
PR 24-FEB-2000; 2000US-0184564P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189374P.
PR 17-MAR-2000; 2000US-0190376P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205315P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225467P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-022824P.
PR 01-SEP-2000; 2000US-0225287P.
PR 01-SEP-2000; 2000US-0225343P.
PR 01-SEP-2000; 2000US-0225344P.
PR 01-SEP-2000; 2000US-0225345P.
PR 05-SEP-2000; 2000US-0225509P.
PR 05-SEP-2000; 2000US-0225513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0235802P.
PR 02-OCT-2000; 2000US-0237037P.

PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239353P.
PR 13-OCT-2000; 2000US-0239357P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249246P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249266P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251983P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and metastasis.
Disclosure; SEQ ID No 34611; 3071pp + Sequence Listing; English.

XX
CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting the
CC nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention
XX
SQ Sequence 21436 BP; 5267 A; 4950 C; 5418 G; 5801 T; 0 U; 0 Other;

Query Match 29.9%; Score 722; DB 4; Length 21436;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1457 AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACAGCGCCAGGTTCCATGT 1516
DB 9454 AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACAGCGCCAGGTTCCATGT 9513
QY 1517 CGAGTCCTAACGGGAATCCAGTGTCAACAGGGCTAGTCTGTATGCTTCAGCTTCTG 1576
DB 9514 CGAGTCCTAACGGGAATCCAGTGTCAACAGGGCTAGTCTGTATGCTTCAGCTTCTG 9573
QY 1577 TCCAAACCTGGTCAGTTAATAGTCTCTCAGTGTATCCAGACATCTTAAGAACCAAAAA 1636
DB 9574 TCCAAACCTGGTCAGTTAATAGTCTCTCAGTGTATCCAGACATCTTAAGAACCAAAAA 9633
QY 1637 TTACAATCAGTATTCAACAAGTTGCCCTGTTCCGACGTTGTCAGTCTCAACCTTACCTTC 1696
DB 9634 TTACAATCAGTATTCAACAAGTTGCCCTGTTCCGACGTTGTCAGTCTCAACCTTACCTTC 9693
QY 1697 ATAGTAATCTTTGAGAACCTTACCAAGCCGTTCCCTCTTCTTCACTTACCAATCTTG 1756
DB 9694 ATAGTAATCTTTGAGAACCTTACCAAGCCGTTCCCTCTTCTTCACTTACCAATCTTG 9753
QY 1757 CAGTACAGTCTACCTCGAAACGATCTACGATGTCACTAGTCTTCTAGTAAAGTAACAAACCGA 1816
DB 9754 CAGTACAGTCTACCTCGAAACGATCTACGATGTCACTAGTCTTCTAGTAAAGTAACAAACCGA 9813
QY 1817 TCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGATGAATGGCAATFCCAAGCTGAATCCA 1876
DB 9814 TCCCCCGCAGTGAATCTCTGCTCCAGCCCGTGATGAATGGCAATFCCAAGCTGAATCCA 9873
QY 1877 GCGTGTGCTGCTCCCTATGGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGC 1936
DB 9874 GCGTGTGCTGCTCCCTATGGCCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGC 9933
QY 1937 TGGGCAAGGAGAATGGGATTTGGTACGATTTGAGCTCCCACTCTCCCGGCCAAGATGCC 1996
DB 9934 TGGGCAAGGAGAATGGGATTTGGTACGATTTGAGCTCCCACTCTCCCGGCCAAGATGCC 9993
QY 1997 AAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAACGGTCTAATA 2056
DB 9994 AAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCTTAAACGGTCTAATA 10053
QY 2057 GTGCAGACAGCAGTGAACCCGAAAGAAACGCGCTTAGCGCTGATGTCGACGCTGCC 2116
DB 10054 GTGCAGACAGCAGTGAACCCGAAAGAAACGCGCTTAGCGCTGATGTCGACGCTGCC 10113
QY 2117 AAGGCCAGCTCCCTGCTCACTCAGAAAAATCCCTTTGCTAGGGCAAAACGGTCTTCTCGTGA 2176
DB 10114 AAGGCCAGCTCCCTGCTCACTCAGAAAAATCCCTTTGCTAGGGCAAAACGGTCTTCTCGTGA 10173

Wed Aug 18 13:52:58 2004

us-10-049-745-31.oli30.rng

```
QV      2177 AG 2178
Db      10174 AG 10175

RESULT 11
AAK70011
ID AAK70011 standard; DNA; 21436 BP.
XX AC AAK70011;
XX XX
XX DT 06-NOV-2001 (first entry)
XX DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:24823.
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX OS Homo sapiens.
XX XX
XX PN WO200157182-A2.
XX PD 09-AUG-2001.
XX XX
XX PF 17-JAN-2001; 2001WO-US001354.
XX XX
XX PR 31-JAN-2000; 2000US-0179065P.
XX PR 04-FEB-2000; 2000US-0180628P.
XX PR 24-FEB-2000; 2000US-0184664P.
XX PR 02-MAR-2000; 2000US-0186350P.
XX PR 16-MAR-2000; 2000US-0189874P.
XX PR 17-MAR-2000; 2000US-0190076P.
XX PR 18-APR-2000; 2000US-0198123P.
XX PR 19-MAY-2000; 2000US-0205515P.
XX PR 07-JUN-2000; 2000US-0209467P.
XX PR 28-JUN-2000; 2000US-0214886P.
XX PR 30-JUN-2000; 2000US-0215135P.
XX PR 07-JUL-2000; 2000US-0216647P.
XX PR 07-JUL-2000; 2000US-0216880P.
XX PR 11-JUL-2000; 2000US-0217487P.
XX PR 11-JUL-2000; 2000US-0217496P.
XX PR 14-JUL-2000; 2000US-0218290P.
XX PR 26-JUL-2000; 2000US-0220963P.
XX PR 26-JUL-2000; 2000US-0220964P.
XX PR 14-AUG-2000; 2000US-0224518P.
XX PR 14-AUG-2000; 2000US-0224519P.
XX PR 14-AUG-2000; 2000US-0225213P.
XX PR 14-AUG-2000; 2000US-0225214P.
XX PR 14-AUG-2000; 2000US-0225266P.
XX PR 14-AUG-2000; 2000US-0225267P.
XX PR 14-AUG-2000; 2000US-0225268P.
XX PR 14-AUG-2000; 2000US-0225270P.
XX PR 14-AUG-2000; 2000US-0225447P.
XX PR 14-AUG-2000; 2000US-0225757P.
XX PR 14-AUG-2000; 2000US-0225758P.
XX PR 14-AUG-2000; 2000US-0225759P.
XX PR 18-AUG-2000; 2000US-0226279P.
XX PR 22-AUG-2000; 2000US-0226681P.
XX PR 22-AUG-2000; 2000US-0226686P.
XX PR 22-AUG-2000; 2000US-0227182P.
XX PR 23-AUG-2000; 2000US-0227003P.
XX PR 30-AUG-2000; 2000US-0228924P.
XX PR 01-SEP-2000; 2000US-0229287P.
XX PR 01-SEP-2000; 2000US-0229343P.
XX PR 01-SEP-2000; 2000US-0229344P.
XX PR 01-SEP-2000; 2000US-0229345P.
XX PR 05-SEP-2000; 2000US-0229503P.
XX PR 05-SEP-2000; 2000US-0229513P.
XX PR 06-SEP-2000; 2000US-0230437P.
XX PR 06-SEP-2000; 2000US-0230438P.
XX PR 08-SEP-2000; 2000US-0231242P.
XX PR 08-SEP-2000; 2000US-0231243P.
XX PR 08-SEP-2000; 2000US-0231244P.
XX PR 08-SEP-2000; 2000US-0231413P.
XX PR 08-SEP-2000; 2000US-0231414P.
XX PR 08-SEP-2000; 2000US-0232380P.
XX PR 08-SEP-2000; 2000US-0232381P.
XX PR 12-SEP-2000; 2000US-0232397P.
XX PR 14-SEP-2000; 2000US-0232398P.
XX PR 14-SEP-2000; 2000US-0232399P.
XX PR 14-SEP-2000; 2000US-0232400P.
XX PR 14-SEP-2000; 2000US-0232401P.
XX PR 14-SEP-2000; 2000US-0233063P.
XX PR 14-SEP-2000; 2000US-0233064P.
XX PR 14-SEP-2000; 2000US-0233065P.
XX PR 21-SEP-2000; 2000US-0234223P.
XX PR 21-SEP-2000; 2000US-0234274P.
XX PR 25-SEP-2000; 2000US-0234997P.
XX PR 25-SEP-2000; 2000US-0234998P.
XX PR 26-SEP-2000; 2000US-0235484P.
XX PR 27-SEP-2000; 2000US-0235834P.
XX PR 27-SEP-2000; 2000US-0235836P.
XX PR 29-SEP-2000; 2000US-0236327P.
XX PR 29-SEP-2000; 2000US-0236367P.
XX PR 29-SEP-2000; 2000US-0236368P.
XX PR 29-SEP-2000; 2000US-0236369P.
XX PR 29-SEP-2000; 2000US-0236370P.
XX PR 02-OCT-2000; 2000US-0236802P.
XX PR 02-OCT-2000; 2000US-0237037P.
XX PR 02-OCT-2000; 2000US-0237038P.
XX PR 02-OCT-2000; 2000US-0237039P.
XX PR 02-OCT-2000; 2000US-0237040P.
XX PR 13-OCT-2000; 2000US-0239335P.
XX PR 13-OCT-2000; 2000US-0239337P.
XX PR 20-OCT-2000; 2000US-0240960P.
XX PR 20-OCT-2000; 2000US-0241221P.
XX PR 20-OCT-2000; 2000US-0241785P.
XX PR 20-OCT-2000; 2000US-0241786P.
XX PR 20-OCT-2000; 2000US-0241787P.
XX PR 20-OCT-2000; 2000US-0241808P.
XX PR 20-OCT-2000; 2000US-0241809P.
XX PR 20-OCT-2000; 2000US-0241826P.
XX PR 01-NOV-2000; 2000US-0244617P.
XX PR 08-NOV-2000; 2000US-0246474P.
XX PR 08-NOV-2000; 2000US-0246475P.
XX PR 08-NOV-2000; 2000US-0246476P.
XX PR 08-NOV-2000; 2000US-0246477P.
XX PR 08-NOV-2000; 2000US-0246478P.
XX PR 08-NOV-2000; 2000US-0246523P.
XX PR 08-NOV-2000; 2000US-0246524P.
XX PR 08-NOV-2000; 2000US-0246525P.
XX PR 08-NOV-2000; 2000US-0246526P.
XX PR 08-NOV-2000; 2000US-0246527P.
XX PR 08-NOV-2000; 2000US-0246528P.
XX PR 08-NOV-2000; 2000US-0246529P.
XX PR 08-NOV-2000; 2000US-0246610P.
XX PR 08-NOV-2000; 2000US-0246611P.
XX PR 08-NOV-2000; 2000US-0246613P.
XX PR 17-NOV-2000; 2000US-0249207P.
XX PR 17-NOV-2000; 2000US-0249208P.
XX PR 17-NOV-2000; 2000US-0249209P.
XX PR 17-NOV-2000; 2000US-0249210P.
XX PR 17-NOV-2000; 2000US-0249211P.
XX PR 17-NOV-2000; 2000US-0249212P.
XX PR 17-NOV-2000; 2000US-0249213P.
XX PR 17-NOV-2000; 2000US-0249214P.
XX PR 17-NOV-2000; 2000US-0249215P.
XX PR 17-NOV-2000; 2000US-0249216P.
XX PR 17-NOV-2000; 2000US-0249217P.
XX PR 17-NOV-2000; 2000US-0249218P.
XX PR 17-NOV-2000; 2000US-0249244P.
XX PR 17-NOV-2000; 2000US-0249245P.
XX PR 17-NOV-2000; 2000US-0249264P.
XX PR 17-NOV-2000; 2000US-0249265P.
```


us-10-049-745-31.oli30.rng

Wed Aug 18 13:52:58 2004

PR	26-JUL-2000;	2000US-0220963P.	PR	08-NOV-2000;	2000US-0246475P.
PR	26-JUL-2000;	2000US-0220964P.	PR	08-NOV-2000;	2000US-0246476P.
PR	14-AUG-2000;	2000US-0224518P.	PR	08-NOV-2000;	2000US-0246477P.
PR	14-AUG-2000;	2000US-0224519P.	PR	08-NOV-2000;	2000US-0246478P.
PR	14-AUG-2000;	2000US-0225213P.	PR	08-NOV-2000;	2000US-0246523P.
PR	14-AUG-2000;	2000US-0225214P.	PR	08-NOV-2000;	2000US-0246524P.
PR	14-AUG-2000;	2000US-0225266P.	PR	08-NOV-2000;	2000US-0246525P.
PR	14-AUG-2000;	2000US-0225267P.	PR	08-NOV-2000;	2000US-0246526P.
PR	14-AUG-2000;	2000US-0225268P.	PR	08-NOV-2000;	2000US-0246527P.
PR	14-AUG-2000;	2000US-0225270P.	PR	08-NOV-2000;	2000US-0246528P.
PR	14-AUG-2000;	2000US-0225447P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225757P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225758P.	PR	08-NOV-2000;	2000US-0246532P.
PR	14-AUG-2000;	2000US-0225759P.	PR	08-NOV-2000;	2000US-0246532P.
PR	18-AUG-2000;	2000US-0226279P.	PR	08-NOV-2000;	2000US-0246611P.
PR	22-AUG-2000;	2000US-0226681P.	PR	08-NOV-2000;	2000US-0246613P.
PR	22-AUG-2000;	2000US-0226868P.	PR	17-NOV-2000;	2000US-0249207P.
PR	22-AUG-2000;	2000US-0227182P.	PR	17-NOV-2000;	2000US-0249208P.
PR	23-AUG-2000;	2000US-0227009P.	PR	17-NOV-2000;	2000US-0249209P.
PR	30-AUG-2000;	2000US-0228924P.	PR	17-NOV-2000;	2000US-0249210P.
PR	01-SEP-2000;	2000US-0229287P.	PR	17-NOV-2000;	2000US-0249211P.
PR	01-SEP-2000;	2000US-0229343P.	PR	17-NOV-2000;	2000US-0249212P.
PR	01-SEP-2000;	2000US-0229344P.	PR	17-NOV-2000;	2000US-0249213P.
PR	01-SEP-2000;	2000US-0229345P.	PR	17-NOV-2000;	2000US-0249214P.
PR	05-SEP-2000;	2000US-0229509P.	PR	17-NOV-2000;	2000US-0249215P.
PR	05-SEP-2000;	2000US-0229513P.	PR	17-NOV-2000;	2000US-0249216P.
PR	06-SEP-2000;	2000US-0230437P.	PR	17-NOV-2000;	2000US-0249217P.
PR	06-SEP-2000;	2000US-0230438P.	PR	17-NOV-2000;	2000US-0249218P.
PR	08-SEP-2000;	2000US-0231242P.	PR	17-NOV-2000;	2000US-0249244P.
PR	08-SEP-2000;	2000US-0231243P.	PR	17-NOV-2000;	2000US-0249245P.
PR	08-SEP-2000;	2000US-0231244P.	PR	17-NOV-2000;	2000US-0249264P.
PR	08-SEP-2000;	2000US-0231411P.	PR	17-NOV-2000;	2000US-0249265P.
PR	08-SEP-2000;	2000US-0231412P.	PR	17-NOV-2000;	2000US-0249297P.
PR	08-SEP-2000;	2000US-0232080P.	PR	17-NOV-2000;	2000US-0249299P.
PR	08-SEP-2000;	2000US-0232081P.	PR	17-NOV-2000;	2000US-0249300P.
PR	12-SEP-2000;	2000US-0231968P.	PR	01-DEC-2000;	2000US-0250160P.
PR	14-SEP-2000;	2000US-0232397P.	PR	05-DEC-2000;	2000US-0250391P.
PR	14-SEP-2000;	2000US-0232398P.	PR	05-DEC-2000;	2000US-0251030P.
PR	14-SEP-2000;	2000US-0232399P.	PR	05-DEC-2000;	2000US-0251988P.
PR	14-SEP-2000;	2000US-0232400P.	PR	05-DEC-2000;	2000US-0256719P.
PR	14-SEP-2000;	2000US-0232401P.	PR	06-DEC-2000;	2000US-0251479P.
PR	14-SEP-2000;	2000US-0233063P.	PR	08-DEC-2000;	2000US-0251856P.
PR	14-SEP-2000;	2000US-0233064P.	PR	08-DEC-2000;	2000US-0251868P.
PR	14-SEP-2000;	2000US-0233065P.	PR	08-DEC-2000;	2000US-0251869P.
PR	21-SEP-2000;	2000US-0234223P.	PR	08-DEC-2000;	2000US-0251989P.
PR	21-SEP-2000;	2000US-0234274P.	PR	11-DEC-2000;	2000US-0254097P.
PR	25-SEP-2000;	2000US-0234997P.	PR	05-JAN-2001;	2001US-0259678P.
PR	25-SEP-2000;	2000US-0234998P.	XX		
PR	26-SEP-2000;	2000US-0235484P.	PA	(HUMA-) HUMAN GENOME SC: INC.	
PR	27-SEP-2000;	2000US-0235834P.	Rosen CA, Barash SC, Ruben SM;		
PR	27-SEP-2000;	2000US-0235836P.	WPI; 2001-465566/50.		
PR	29-SEP-2000;	2000US-0236327P.	P-PSDB; AAU23207.		
PR	29-SEP-2000;	2000US-0236367P.			
PR	29-SEP-2000;	2000US-0236368P.			
PR	29-SEP-2000;	2000US-0236369P.			
PR	29-SEP-2000;	2000US-0236370P.			
PR	02-OCT-2000;	2000US-0236802P.			
PR	02-OCT-2000;	2000US-0237037P.			
PR	02-OCT-2000;	2000US-0237038P.			
PR	02-OCT-2000;	2000US-0237039P.			
PR	02-OCT-2000;	2000US-0237040P.			
PR	13-OCT-2000;	2000US-0239933P.			
PR	13-OCT-2000;	2000US-0239937P.			
PR	20-OCT-2000;	2000US-0240960P.			
PR	20-OCT-2000;	2000US-0241221P.			
PR	20-OCT-2000;	2000US-0241785P.			
PR	20-OCT-2000;	2000US-0241786P.			
PR	20-OCT-2000;	2000US-0241787P.			
PR	20-OCT-2000;	2000US-0241808P.			
PR	20-OCT-2000;	2000US-0241809P.			
PR	20-OCT-2000;	2000US-0241826P.			
PR	01-NOV-2000;	2000US-0244617P.			
PR	08-NOV-2000;	2000US-0246474P.			

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS), autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and

PF	17-JAN-2001;	2001WO-US00012339;
XX	31-JAN-2000;	2000US-01790655PR
PR	04-FEB-2000;	2000US-01790652PR
PR	24-FEB-2000;	2000US-01845644PR
PR	02-MAR-2000;	2000US-01863504PR
PR	16-MAR-2000;	2000US-01898744PR
PR	17-APR-2000;	2000US-01900767PR
PR	18-APR-2000;	2000US-01981233PR
PR	19-MAY-2000;	2000US-02055151PR
PR	07-JUN-2000;	2000US-02094674PR
PR	28-JUN-2000;	2000US-02148867PR
PR	07-JUL-2000;	2000US-02151353PR
PR	07-JUL-2000;	2000US-02166474PR
PR	07-JUL-2000;	2000US-02168800PR
PR	11-JUL-2000;	2000US-02174874PR
PR	11-JUL-2000;	2000US-02174967PR
PR	14-JUL-2000;	2000US-02182904PR
PR	26-JUL-2000;	2000US-02182963PR
PR	26-JUL-2000;	2000US-02209644PR
PR	14-AUG-2000;	2000US-02245184PR
PR	14-AUG-2000;	2000US-02245194PR
PR	14-AUG-2000;	2000US-02252134PR
PR	14-AUG-2000;	2000US-02252144PR
PR	14-AUG-2000;	2000US-02252667PR
PR	14-AUG-2000;	2000US-02252677PR
PR	14-AUG-2000;	2000US-02252684PR
PR	14-AUG-2000;	2000US-02252704PR
PR	14-AUG-2000;	2000US-02254474PR
PR	14-AUG-2000;	2000US-02255774PR
PR	14-AUG-2000;	2000US-02257584PR
PR	18-AUG-2000;	2000US-02257594PR
PR	18-AUG-2000;	2000US-02262719PR
PR	22-AUG-2000;	2000US-02266681PR
PR	22-AUG-2000;	2000US-02268684PR
PR	22-AUG-2000;	2000US-02271934PR
PR	23-AUG-2000;	2000US-02270094PR
PR	30-AUG-2000;	2000US-02288924PR
PR	01-SEP-2000;	2000US-02292874PR
PR	01-SEP-2000;	2000US-02293434PR
PR	01-SEP-2000;	2000US-02293444PR
PR	01-SEP-2000;	2000US-02293454PR
PR	05-SEP-2000;	2000US-02295094PR
PR	05-SEP-2000;	2000US-02295134PR
PR	06-SEP-2000;	2000US-02304374PR
PR	06-SEP-2000;	2000US-02304384PR
PR	08-SEP-2000;	2000US-02312424PR
PR	08-SEP-2000;	2000US-02312434PR
PR	08-SEP-2000;	2000US-02312444PR
PR	08-SEP-2000;	2000US-02314134PR
PR	08-SEP-2000;	2000US-02314144PR
PR	08-SEP-2000;	2000US-02320804PR
PR	12-SEP-2000;	2000US-02321968PR
PR	14-SEP-2000;	2000US-02323974PR
PR	14-SEP-2000;	2000US-02323984PR
PR	14-SEP-2000;	2000US-02323994PR
PR	14-SEP-2000;	2000US-02324004PR
PR	14-SEP-2000;	2000US-02324014PR
PR	14-SEP-2000;	2000US-02330634PR
PR	14-SEP-2000;	2000US-02330644PR
PR	14-SEP-2000;	2000US-02330654PR
PR	21-SEP-2000;	2000US-02342234PR
PR	21-SEP-2000;	2000US-02342244PR
PR	25-SEP-2000;	2000US-02349974PR
PR	25-SEP-2000;	2000US-02349984PR
PR	26-SEP-2000;	2000US-02354844PR
PR	27-SEP-2000;	2000US-02356834PR
PR	27-SEP-2000;	2000US-02356844PR
PR	29-SEP-2000;	2000US-02363274PR
PR	29-SEP-2000;	2000US-02363284PR
PR	29-SEP-2000;	2000US-02363674PR
PR	29-SEP-2000;	2000US-02363684PR
PR	29-SEP-2000;	2000US-02363694PR

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases.

Claim 4; SEQ ID NO 839; 1180pp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS), autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. AAU40785-AAU41684 represent cDNA sequences encoding for the novel human enzyme polypeptides of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 577 BP; 138 A; 138 C; 147 G; 151 T; 0 U; 3 Other;

Query Match 17.6%; Score 424; DB 4; Length 577;
Best Local Similarity 100.0%; Pred. No. 3 4e-197;
Matches 424; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	35	GGGATGGAGCGAGCGCGGCTCAGAGTTGAAAGAACCATAGTTGACAAAGCTT	94
DB	46	GGGATGGAGCGAGCGCGGCTCAGAGTTGAAAGAACCATAGTTGACAAAGCTT	105
QY	95	CTGAATCTTCAGACCCATCAGCCCTATCAGATCAGCCTCGGAGTCCGAGCAGTCTCAC	154
DB	106	CTGAATCTTCAGACCCATCAGCCCTATCAGATCAGCCTCGGAGTCCGAGCAGTCTCAC	165
QY	155	CTGGAGACATGGATGAGTCTTCCAGCTGGGTGCTGTCTTCAATTGAATGATGT	214
DB	166	CTGGAGACATGGATGAGTCTTCCAGCTGGGTGCTGTCTTCAATTGAATGATGT	225
QY	215	CAAAATCACACATCTTCTTTAGGACAGTACTCTGGTGTCTGTAGTTATTCGAGTTTCATCTG	274
DB	226	CAAAATCACACATCTTCTTTAGGACAGTACTCTGGTGTCTGTAGTTATTCGAGTTTCATCTG	285
QY	275	TACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTC	334
DB	286	TACCTGATAAATCAAAACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTC	345
QY	335	CTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTGTCTTAAGTGGCAACAACTCATTA	394
DB	346	CTCCACAGAAAGTTCTTTTCCCATCTGAGAGATTGTCTTAAGTGGCAACAACTCATTA	405
QY	395	CAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATCAGCAGTGCAGT	454
DB	406	GAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATCAGCAGTGCAGT	465
QY	455	GTTT 458	
DB	466	GTTT 469	

RESULT 14
AA192171
ID AA192171 standard; cDNA; 411 BP.
XX
AC AA192171;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12231.

29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX PA
XX PI Rosen CA, Barash SC, Ruben SM;
XX XX
XX WPI; 2001-465566/50.
XX P-PSDB; AAU23743.
XX

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
OS
XX WO200164835-A2.
PN
XX
PD
XX
XX 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
XX (HYSE-) HYSEQ INC.
PA
PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI: 2001-514838/56.
DR P-PSDB; AA012240.
DR
XX
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
PT
XX
PS Claim 1; SEQ ID NO 12231; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 411 BP; 108 A; 99 C; 99 G; 105 T; 0 U; 0 Other;

Query Match 17.0%; Score 411; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 8.3e-191;
Matches 411; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 36 GGATGGAGCGAGCGCGGTCAGAGTTGAACAATGACCAATGACCAAGCTTC 95
DB 1 GGATGGAGCGAGCGCGGTCAGAGTTGAACAATGACCAATGACCAAGCTTC 60
QY 96 TGAATCTTCAGACCCATCAGCCTATCAGATCAGCTGGCAGCTCCGAGGAGCTCACC 155
DB 61 TGAATCTTCAGACCCATCAGCCTATCAGATCAGCTGGCAGCTCCGAGGAGCTCACC 120
QY 156 TGGAGAGATGATCAGGTTCTCCAGCTGGGGTCTGTCTTCATGATGATGATGTC 215
DB 121 TGGAGAGATGATCAGGTTCTCCAGCTGGGGTCTGTCTTCATGATGATGATGTC 180
QY 216 AAATCACACACTTCTTTAGGACAGTACCTGGTGTGTAGTTTATTCGAGTTCTCTGT 275
DB 181 AAATCACACACTTCTTTAGGACAGTACCTGGTGTGTAGTTTATTCGAGTTCTCTGT 240
QY 276 ACTGTATAATCAAAACATCACCACAAAAGGATCAAGCCCTAGGTGATGCGTCCTCC 335
DB 241 ACTGTATAATCAAAACATCACCACAAAAGGATCAAGCCCTAGGTGATGCGTCCTCC 300
QY 336 TCCACAGAAAGTTCTTTTCCATCTGAGAGATTTGCTTAAGTGGCAACAACTCATAG 395
DB 301 TCCACAGAAAGTTCTTTTCCATCTGAGAGATTTGCTTAAGTGGCAACAACTCATAG 360

QY 396 AGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGAGC 446
DB 361 AGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGAGC 411
RESULT 15
ABK30453
ID ABK30453 standard; cDNA; 722 BP.
XX
AC ABK30453;
XX
DT 23-APR-2002 (first entry)
XX
DE Human G-protein-coupled protease #223.
XX
KW Human; ss; gene; G-protein-coupled protease; gene therapy; transgenic;
KW protease mediated disorder; proliferative disorder;
KW differentiative disorder; developmental disorder;
KW haematopoietic disorder.
XX
OS Homo sapiens.
XX
PN US6331427-B1.
XX
XX 18-DEC-2001.
PD
XX
PF 26-MAR-1999; 99US-00280116.
PF
PR 26-MAR-1999; 99US-00280116.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Robison KE;
XX
XX WPI: 2002-129545/17.
XX
XX New polynucleotides encoding protease homologs of the G-protein-coupled
PT protease family, useful in identifying agonists and antagonists for
PT diagnosis and treatment of protease mediated disorders.
XX
PS Disclosure; Col 251-252; 246pp; English.
XX
CC The invention relates to an isolated human protease nucleic acid molecule
CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
CC defined in the specification. Also disclosed are production of an
CC isolated polypeptide encoded by the nucleic acid, comprising introducing
CC the nucleic acid into a host cell and culturing under conditions to
CC express the protein from the nucleic acid, use of an antibody to detect
CC the encoded protein in a sample and to modulate its in vivo activity,
CC identifying agents that bind to the protein and identification of a
CC polynucleotide agent that modulates the expression of the nucleic acid or
CC its complement (i.e. gene therapy). The nucleic acid can be used to
CC identify an agent that modulates the expression or activity of the
CC nucleic acid, and can be used to isolate the protein. The nucleic acid
CC can be used in diagnostic assays for determining nucleic acid expression
CC as well as activity in the context of a biological sample (e.g., blood,
CC serum, cells, tissue) to determine whether an individual has a disease or
CC disorder, or is at risk of developing a disease or disorder, associated
CC with aberrant expression or activity of the nucleic acid. The nucleic
CC acid can be used to detect mutations in protease genes and gene
CC expression products such as mRNA. The nucleic acid can be used as
CC hybridisation probes to detect naturally-occurring genetic mutations in a
CC protease gene. The nucleic acid can be used in drug screening methods to
CC identify agonists and antagonists that can be used to diagnose and treat
CC such protease mediated disorders e.g., proliferative, differentiative,
CC developmental or haematopoietic disorders. The nucleic acid can be used
CC as probes, primers, in biological assays, to determine patterns of gene
CC expression, to design ribozymes and to construct transgenic animals. The
CC present sequence represents one of the 268 disclosed human G-protein-
CC coupled protease cDNA sequences
XX
SQ Sequence 722 BP; 189 A; 186 C; 145 G; 196 T; 0 U; 6 Other;

Query Match 14.8%; Score 357; DB 6; Length 722;
Best Local Similarity 100.0%; Pred. No. 2.5e-164; Indels 0; Gaps 0;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1061 AGGATGTAATACCTGAGTATCTTGATATTCGGCCATATATGCTCAACCCCAACGGAG 1120
Db 107 AGGATGTGAATACCTGAGTATCTTGATATTCGGCCATATATGCTCAACCCCAACGGAG 166
QY 1121 AGCCAAATGCTACGCTCTTGATGAGTCTGCTCCACACTGGTTTAAATGCCATGCTG 1180
Db 167 AGCCAAATGCTACGCTCTTGATGAGTCTGCTCCACACTGGTTTAAATGCCATGCTG 226
QY 1181 GCATTAATCTGCTACATAAAGCTAGCAATGGCTCTGGTATCAATGAATGACTCCA 1240
Db 227 GCATTAATCTGCTACATAAAGCTAGCAATGGCTCTGGTATCAATGAATGACTCCA 286
QY 1241 TTGTATCTACCACTGATATTTAGATCGGTACTCAGCCCAACAGGCTATGTCTTTTATA 1300
Db 287 TTGTATCTACCACTGATATTTAGATCGGTACTCAGCCCAACAGGCTATGTCTTTTATA 346
QY 1301 TCAGGTCCCATGATGTGAAAATGAGGTGAATTAATCCACCCATAGCCCGGCC 1360
Db 347 TCAGGTCCCATGATGTGAAAATGAGGTGAATTAATCCACCCATAGCCCGGCC 406
QY 1361 AGTCTCTCCCCCGCTCATGCTCAGCGGTTGTCCACCAACAGGCTGCGC 1417
Db 407 AGTCTCTCCCCCGCTCATGCTCAGCGGTTGTCCACCAACAGGCTGCGC 463

RESULT 16
AAI22307/c
ID AAI22307 standard; DNA; 333 BP.
XX AC AAI22307;
XX DT 12-OCT-2001 (first entry)
XX DE Probe #12240 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human cervical epithelial cells.
XX PS Claim 25; SEQ ID NO 12240; 487pp; English.

The present invention relates to human single exon nucleic acid probes (SENP). The present sequence is one such probe. The SENPs are derived from human HeLa cells. The SENPs can be used to produce a single exon microarray, which can be used for measuring human gene expression in a

CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
QY Query Match 13.8%; Score 333; DB 4; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152; Indels 0; Gaps 0;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 CCCGTGATGAATGGCAATCCAAAGCTGAACTCCAGCGTCTGGTGCCTATGGCGCCGAG 1902
Db 333 CCCGTGATGAATGGCAATCCAAAGCTGAACTCCAGCGTCTGGTGCCTATGGCGCCGAG 274
QY 1903 TCTCTGAGAGCTCTTACGAGAGTCAAAAGGGCTGGCAAGGAGATGGATTGTTACG 1962
Db 273 TCTCTGAGAGCTCTTACGAGAGTCAAAAGGGCTGGCAAGGAGATGGATTGTTACG 214
QY 1963 ATTGTGAGCTCCACACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACGAG 2022
Db 213 ATTGTGAGCTCCACACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCCACTCCGACGAG 154
QY 2023 CTTCAAGAAACCCATGACCCCTTAAACCGTGTCTAATAGTGCAGACAGCAGTACCCGAAA 2082
Db 153 CTTCAAGAAACCCATGACCCCTTAAACCGTGTCTAATAGTGCAGACAGCAGTACCCGAAA 94
QY 2083 GAAACGGCGCTAGCGCTGATGTGTCAGTGCAGTGCAGGCGCAGCTGCCCTGCACTCAGAA 2142
Db 93 GAAACGGCGCTAGCGCTGATGTGTCAGTGCAGTGCAGGCGCAGCTGCCCTGCACTCAGAA 34
QY 2143 AATCCCTTTGCTAAGCAACCGTCTTCTCTGGA 2175
Db 33 AATCCCTTTGCTAAGCAACCGTCTTCTCTGGA 1

RESULT 17
ABA67385/c
ID ABA67385 standard; DNA; 333 BP.
XX AC ABA67385;
XX DT 01-FEB-2002 (first entry)
XX DE Human foetal liver single exon nucleic acid probe #15690.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX DR Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

PT gene expression in human fetal liver.

PS Claim 4; SEQ ID NO 15690; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human fetal liver. The
CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;

Query Match 13.8%; Score 333; DB 4; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1843 CCCGTGATGAATGGCAAAATCCAAAGCTGAATCCAGCGTGTCTGTGCTCCCTATGCGCCGAG 1902

DB 333 CCCGTGATGAATGGCAAAATCCAAAGCTGAATCCAGCGTGTCTGTGCTCCCTATGCGCCGAG 274

QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAAAGGGCTGGGCAAGAGATGGGATTGGTACG 1962

DB 273 TCCTCTGAGGACTCTGACGAGGAGTCAAAAGGGCTGGGCAAGAGATGGGATTGGTACG 214

QY 1963 ATTGTGAGTCTCCACTCTCCCGGCCAAGATGCGGAGATGAGAGGCCACTCGCACGAG 2022

DB 213 ATTGTGAGTCTCCACTCTCCCGGCCAAGATGCGGAGATGAGAGGCCACTCGCACGAG 154

QY 2023 CTTCAAGAACCCATGACCTCAACCTTAACCGTGCTAATAGTGACAGAGGACAGTCCCGGAA 2082

DB 153 CTTCAAGAACCCATGACCTCAACCGTGCTAATAGTGACAGAGGACAGTCCCGGAA 94

QY 2083 GAAACGGCTTAGCGCTGAGGTGCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 2142

DB 93 GAAACGGCTTAGCGCTGAGGTGCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 34

QY 2143 AATCCCTTTGCTTAAGGCAACGGTCTTCTGGA 2175

DB 33 AATCCCTTTGCTTAAGGCAACGGTCTTCTGGA 1

RESULT 18

ABA49470/c

ID ABA49470 standard; DNA; 333 BP.

XX ABA49470;

AC ABA49470;

XX 01-FEB-2002 (first entry)

XX Human breast cell single exon nucleic acid probe #8165.

XX Human; microarray; single exon probe; gene expression; breast; disease;

XX cancer; ss.

XX Homo sapiens.

XX WO200157271-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000662.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-496933/54.

XX New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.

XX Claim 4; SEQ ID NO 8165; 327pp + Sequence Listing; English.

XX The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of the human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or
CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;

Query Match 13.8%; Score 333; DB 4; Length 333;

Best Local Similarity 100.0%; Pred. No. 1.5e-152;

Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1843 CCCGTGATGAATGGCAAAATCCAAAGCTGAATCCAGCGTGTCTGTGCTCCCTATGCGCCGAG 1902

DB 333 CCCGTGATGAATGGCAAAATCCAAAGCTGAATCCAGCGTGTCTGTGCTCCCTATGCGCCGAG 274

QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAAAGGGCTGGGCAAGAGATGGGATTGGTACG 1962

DB 273 TCCTCTGAGGACTCTGACGAGGAGTCAAAAGGGCTGGGCAAGAGATGGGATTGGTACG 214

QY 1963 ATTGTGAGTCTCCACTCTCCCGGCCAAGATGCGGAGATGAGAGGCCACTCGCACGAG 2022

DB 213 ATTGTGAGTCTCCACTCTCCCGGCCAAGATGCGGAGATGAGAGGCCACTCGCACGAG 154

QY 2023 CTTCAAGAACCCATGACCTCAACCGTGCTAATAGTGACAGAGGACAGTCCCGGAA 2082

DB 153 CTTCAAGAACCCATGACCTCAACCGTGCTAATAGTGACAGAGGACAGTCCCGGAA 94

QY 2083 GAAACGGCTTAGCGCTGAGGTGCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 2142

DB 93 GAAACGGCTTAGCGCTGAGGTGCCAGCTGCCAAGGCCAGCCCTGCCCTGCACTCAGAA 34

QY 2143 AATCCCTTTGCTTAAGGCAACGGTCTTCTGGA 2175

DB 33 AATCCCTTTGCTTAAGGCAACGGTCTTCTGGA 1

RESULT 19

AAK41552/c

ID AAK41552 standard; DNA; 333 BP.

XX AAK41552;

AC AAK41552;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID NO: 16109.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX

KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX Homo sapiens.
OS WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human bone marrow.
XX
XX Example 4; SEQ ID NO 16109; 658pp + Sequence Listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention
XX
XX SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
Query Match 13.8%; Score 333; DB 4; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 CCGTGATGATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 1902
Db 333 CCGTGATGATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 274
QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 1962
Db 273 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 214
QY 1963 ATTGTGAGTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGACGAG 2022
Db 213 ATTGTGAGTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGACGAG 154
QY 2023 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGCAGTACCCGAAA 2082
Db 153 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGCAGTACCCGAAA 94
QY 2083 GAAACGGCCTAGGCGCTGTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTGACTCAGAA 2142
Db 93 GAAACGGCCTAGGCGCTGTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTGACTCAGAA 34
QY 2143 AATCCCTTTGCTAAGGCAAAACGGTCTTCTGGA 2175
Db 33 AATCCCTTTGCTAAGGCAAAACGGTCTTCTGGA 1
RESULT 20
AAI07994/c
ID AAI07994 standard; DNA; 333 BP.
XX
AC AAI07994;
XX

DT 09-OCT-2001 (first entry)
XX Probe #7985 used to measure gene expression in human breast sample.
DE
XX
XX Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX WO200157270-A2.
XX 09-AUG-2001.
XX 29-JAN-2001; 2001WO-US000661.
XX 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX Novel single exon nucleic acid probe used to measuring gene expression in
PT a human breast.
PT
XX Claim 25; SEQ ID NO 7985; 322pp; English.
PS
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and non-
CC carcinoma tumours. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;
Query Match 13.8%; Score 333; DB 5; Length 333;
Best Local Similarity 100.0%; Pred. No. 1.5e-152;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1843 CCGTGATGATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 1902
Db 333 CCGTGATGATGGCAATCCAACTGAACTCCAGCGTGTGGTCCCTATGGCGCCGAG 274
QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 1962
Db 273 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGGAGATGGGATTGGTACG 214
QY 1963 ATTGTGAGTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGACGAG 2022
Db 213 ATTGTGAGTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGACGAG 154
QY 2023 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGCAGTACCCGAAA 2082
Db 153 CTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCAGACAGCAGCAGTACCCGAAA 94
QY 2083 GAAACGGCCTAGGCGCTGTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTGACTCAGAA 2142
Db 93 GAAACGGCCTAGGCGCTGTGATGGTCCAGCTGCCAAGGCCAGCCTGCCCTGACTCAGAA 34

QY 2143 AATCCCTTTGCTAGGCAACGGTCTTCCTGGA 2175
 Db |||||
 33 AATCCCTTTGCTAGGCAACGGTCTTCCTGGA 1

RESULT 21

ABSL5563/c
 ID ABSL5563 standard; DNA; 333 BP.

XX AC ABSL5563;

XX DT 19-AUG-2002 (first entry)

XX DE Human genome-derived single exon probe ORF from lung SEQ ID No 15554.

XX KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
 KW Chronic obstructive pulmonary disease; interstitial lung disease;
 KW Familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW Tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW Pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW Pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW Primary ciliary dyskinesia; pulmonary hypertension;
 KW Hyaline membrane disease; open reading frame; ORF.

OS Homo sapiens.

XX XX WO200186003-A2.

XX PN 15-NOV-2001.

XX PD 30-JAN-2001; 2001WO-US000665.

XX PF 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX XX WPI; 2002-114183/15.

XX PT Spatially-addressable set of single exon nucleic acid probes, used to
 PT measure gene expression in human lung samples.

XX FS Claim 4; SEQ ID NO 15554; 634pp; English.

XX CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human lung comprising single exon nucleic acid probes having one of
 CC 12614 nucleic acid sequences mentioned in the specification, or their
 CC complements or the 12387 open reading frames derived from the 12614
 CC probes. Also included are a microarray comprising the novel set of probes
 CC ; the novel set of probes which hybridise at high stringency to a nucleic
 CC acid expressed in the human lung; measuring gene expression in a nucleic
 CC derived from human lung, comprising (a) contacting the array with a
 CC collection of detectably labeled nucleic acids derived from human lung
 CC mRNA, and (b) measuring the label detectably bound to each probe of the
 CC array; identifying exons in a eukaryotic genome, comprising (a)
 CC algorithmically predicting at least one exon from genomic sequences of
 CC the eukaryote; and (b) detecting specific hybridisation of detectably
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
 CC having a fragment identical to the predicted exon, the probe is included
 CC in the above mentioned microarray; assigning exons to a single gene,
 CC comprising (a) identifying exons from genomic sequence by the method
 CC above and (b) measuring the expression of each of the exons in several
 CC tissues and/or cell types using hybridisation to a single exon
 CC microarrays having a probe with the exon, where a common pattern of

CC expression of the exons in the tissues and/or cell types indicates that
 CC the exons should be assigned to a single gene; a peptide comprising one
 CC of 12011 sequences, mentioned in the specification, or encoded by the
 CC probes/open reading frames (ORF). The probes are used for gene expression
 CC analysis, and for identifying exons in a gene, particularly using human
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,
 CC Pudlak syndrome, Gaucher's disease, Niemann-Pick disease, Hermansky-
 CC histiocytosis, lymphangioleiomyomatosis, pulmonary haemosiderosis, pulmonary
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The
 CC present sequence is a single exon probe open reading frame of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pt_sequences

SQ Sequence 333 BP; 57 A; 96 C; 95 G; 85 T; 0 U; 0 Other;

Query Match 13.8%; Score 333; DB 6; Length 333;
 Best Local Similarity 100.0%; Pred. No. 1.5e-152;
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1843 CCGTGATGAATGGCAATCCAAAGCTGAATCCAGCGTCTGGTGCCTATGGCGCGAG 1902
 Db |||||
 333 CCGTGATGAATGGCAATCCAAAGCTGAATCCAGCGTCTGGTGCCTATGGCGCGAG 274
 QY 1903 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTGTACG 1962
 Db |||||
 273 TCCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGAGAAATGGGATTGTACG 214
 QY 1963 ATGTGAGCTCCCACTCTCCGGCCCAAGATGCGGAGGAGGCGCACTCGCACGAG 2022
 Db |||||
 213 ATGTGAGCTCCCACTCTCCGGCCCAAGATGCGGAGGAGGCGCACTCGCACGAG 154
 QY 2023 CTTCAAGAACCATGACCCCTAAACGGTGCTAATAGTGACAGACGACAGTACCCGAAA 2082
 Db |||||
 153 CTTCAAGAACCATGACCCCTAAACGGTGCTAATAGTGACAGACGACAGTACCCGAAA 94
 QY 2083 GAAAAAGCGCTAGCGCTGATGGTCCAGCTCCGAGGCGGAGGCGCACTCGCACGAG 2142
 Db |||||
 93 GAAAAAGCGCTAGCGCTGATGGTCCAGCTCCGAGGCGGAGGCGCACTCGCACGAG 34
 QY 2143 AATCCCTTTGCTAAGGCAACGGTCTTCCTGGA 2175
 Db |||||
 33 AATCCCTTTGCTAAGGCAACGGTCTTCCTGGA 1

RESULT 22

AA113071/c

ID AA113071 standard; DNA; 498 BP.

XX AC AA113071;

XX DT 12-OCT-2001 (first entry)

XX DE Probe #3004 for gene expression analysis in human cervical cell sample.
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer; ss.

XX OS Homo sapiens.

XX XX WO200157278-A2.

XX XX 09-AUG-2001.

XX PD 30-JAN-2001; 2001WO-US000670.

XX PF 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00603408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234587P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human fetal liver.

XX Claim 1; SEQ ID NO 3075; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

XX human gene expression in a sample derived from human foetal liver. The

XX single exon nucleic acid probes may be used for predicting, measuring and

XX displaying gene expression in samples derived from human fetal liver. The

XX present sequence is a single exon nucleic acid probe of the invention.

XX Note: The sequence data for this patent did not form part of the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences

XX Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;

XX

Query Match 13.0%; Score 314; DB 4; Length 498;

Best Local Similarity 100.0%; Pred. No. 3.2e-143; Indels 0; Gaps 0;

Matches 314; Conservative 0; Mismatches 0;

Qy 1865 AGCTGAACCTCCAGCTGCTGCTGCTCCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGG 1924

Db 498 AGCTGAACCTCCAGCTGCTGCTGCTCCCTATGCGCGCGAGTCTCTGAGGACTCTGACGAGG 439

Qy 1925 AGTCAAAGGGCGCTGxGCAAGAGAAATGGGATTGGTACGATTGTGAGCTCCCACTCTCCCG 1984

Db 438 AGTCAAAGGGCGCTGxGCAAGAGAAATGGGATTGGTACGATTGTGAGCTCCCACTCTCCCG 379

Qy 1985 GCCAAGATCCGAGAGATGAGAGGCCACTCCGACAGAGCTTCAAGAACCCCATGACCTAA 2044

Db 378 GCCAAGATCCGAGAGATGAGAGGCCACTCCGACAGAGCTTCAAGAACCCCATGACCTAA 319

Qy 2045 ACGGTGCTAATAGTSCAGACGACGACAGTACCCGAAAGAAAACGGCTTAGCCCTGATG 2104

Db 318 ACGGTGCTAATAGTSCAGACGACGACAGTACCCGAAAGAAAACGGCTTAGCCCTGATG 259

Qy 2105 GTGCCAGCTGCCAAAGCGCCAGCTGCCCTGCACTCAGAAATCCCTTTGTGAAGCAACG 2164

Db 258 GTGCCAGCTGCCAAAGCGCCAGCTGCCCTGCACTCAGAAATCCCTTTGTGAAGCAACG 199

Qy 2165 GTCTTCCTGGAAAG 2178

Db 198 GTCTTCCTGGAAAG 185

RESULT 24

ABA44316/c

ID ABA44316 standard; DNA; 498 BP.

XX AC ABA44316;

XX DT 01-FEB-2002 (first entry)

XX Human breast cell single exon nucleic acid probe #3011.

PR	03-AUG-2000; 2000US-00632366.
PR	21-SEP-2000; 2000US-0234687P.
PR	27-SEP-2000; 2000US-0236359P.
PR	04-OCT-2000; 2000GB-00024263.
XX	(MOLE-) MOLECULAR DYNAMICS INC.
XX	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	WPI; 2001-488901/53.
DR	Human genome-derived single exon nucleic acid probes useful for analyzing
PT	gene expression in human cervical epithelial cells.
PT	Human genome-derived single exon nucleic acid probes useful for analyzing
PP	gene expression in human cervical epithelial cells.
XX	Claim 25; SEQ ID NO 3004; 487pp; English.
XX	The present invention relates to human single exon nucleic acid probes
CC	(SENPs). The present sequence is one such probe. The SENPs are derived
CC	from human HeLa cells. The SENPs can be used to produce a single exon
CC	microarray, which can be used for measuring human gene expression in a
CC	sample derived from human cervical epithelial cells. By measuring gene
CC	expression, the probes are therefore useful in grading and/or staging of
CC	diseases of the cervix, notably cervical cancer. Note: The sequence data
CC	for this patent did not form part of the printed specification, but was
CC	obtained in electronic format directly from WIPO at
CC	ftp.wipo.int/pub/published_pct_sequences
XX	Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;
XX	Query Match 13.0%; Score 314; DB 4; Length 498;
XX	Best Local Similarity 100.0%; Pred. No. 3.2e-143;
XX	Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1865 AGCTGAACTCAGGCTCTGGTGCCTATCGGCCGAGTGCTCTGAGGACTCTGACGAGG 1924
DB	498 AGCTGAACTCAGGCTCTGGTGCCTATCGGCCGAGTGCTCTGAGGACTCTGACGAGG 439
QY	1925 AGTCAAAAGGGCTGGGCAAGAGAGATGGGATTGTGTACGATTTGTGAGCTCCCCTCCTCCCG 1984
DB	438 AGTCAAAAGGGCTGGGCAAGAGAGATGGGATTGTGTACGATTTGTGAGCTCCCCTCCTCCCG 379
QY	1985 GCCAAGATGCCGAAGATGAGAGAGCCACTCGCACAGAGTTCAAGAACCATGACCCATA 2044
DB	378 GCCAAGATGCCGAAGATGAGAGAGCCACTCGCACAGAGTTCAAGAACCATGACCCATA 319
QY	2045 ACGGTGCTAATAGTGCAGACAGCAGACAGTGNACCCGAAAGAAAACGGCTTAGCGCTGATG 2104
DB	318 ACGGTGCTAATAGTGCAGACAGCAGACAGTGNACCCGAAAGAAAACGGCTTAGCGCTGATG 259
QY	2105 GTGCCAGCTGCAAGGGCCAGCTGCCCCGTGCACTCAGAAAAATCCCTTTGCTTAAGCAAACG 2164
DB	258 GTGCCAGCTGCAAGGGCCAGCTGCCCCGTGCACTCAGAAAAATCCCTTTGCTTAAGCAAACG 199
QY	2165 GTCTTCCTGGAAAG 2178
DB	198 GTCTTCCTGGAAAG 185
RESULT 23	
ID	ABA54770/c
ID	ABA54770 standard; DNA; 498 BP.
XX	AC ABA54770;
XX	01-FEB-2002 (first entry)
DE	Human foetal liver single exon nucleic acid probe #3075.
XX	Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
OS	Homo sapiens.
PN	WO200157277-A2.

1865 AGCTGAACCTCAGCGTGTCTGTGCGCCTATGCGCGAGTCTCTGAGGACTCTGAGGAGG 1924
438 AGCTGAACCTCAGCGTGTCTGTGCGCCTATGCGCGAGTCTCTGAGGACTCTGAGGAGG 439
1925 AGTCAAGGGGCTGGCGAAGGAGATGGATTGGTACGATTGTGAGTCCCACTCTCCCG 1984
438 AGTCAAGGGGCTGGCGAAGGAGATGGATTGGTACGATTGTGAGTCCCACTCTCCCG 379
1985 GCCAAGATGCCAAGATGAGGAGGCGCACTCCGACAGAGCTTCAAGAACCCATGACCTAA 2044
378 GCCAAGATGCCAAGATGAGGAGGCGCACTCCGACAGAGCTTCAAGAACCCATGACCTAA 319
2045 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGCGCTGATG 2104
318 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGCGCTGATG 259
2105 GTGCCAGCTGCCAAGCGCCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAACG 2164
258 GTGCCAGCTGCCAAGCGCCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAACG 199
2165 GTCTTCTCTGGAAG 2178
198 GTCTTCTCTGGAAG 185
RESULT 27
ABS03008/c
ID ABS03008 standard; DNA; 498 BP.
XX
AC ABS03008;
XX
XX 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe from lung SEQ ID No 2999.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease.
XX
OS Homo sapiens.
XX
XX WO200186003-A2.
XX
XX 15-NOV-2001.
XX
XX 30-JAN-2001; 2001WO-US000665.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 1; SEQ ID NO 2999; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon

2045 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGCGCTGATG 2104
318 ACGGTGCTTAATAGTCAGACAGCGACAGTGAACCGAAGAAACCGCTAGCGCTGATG 259
2105 GTGCCAGCTGCCAAGCGCCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAACG 2164
258 GTGCCAGCTGCCAAGCGCCAGCTGCGCTGCACTCAGAAATCCCTTTGCTAAGGCAACG 199
2165 GTCTTCTCTGGAAG 2178
198 GTCTTCTCTGGAAG 185
RESULT 26
AAI02976/c
ID AAI02976 standard; DNA; 498 BP.
XX
AC AAI02976;
XX
XX 09-OCT-2001 (first entry)
XX
DE Probe #2967 used to measure gene expression in human breast sample.
XX
XX Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
XX WO200157270-A2.
XX
XX 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US000661.
XX
XX 04-FEB-2000; 2000US-0180312P.
XX
XX 26-MAY-2000; 2000US-0207456P.
XX
XX 30-JUN-2000; 2000US-00608408.
XX
XX 03-AUG-2000; 2000US-00632366.
XX
XX 21-SEP-2000; 2000US-0234687P.
XX
XX 27-SEP-2000; 2000US-0236359P.
XX
XX 04-OCT-2000; 2000GB-00024263.
XX
(MOLE-) MOLECULAR DYNAMICS INC.
XX
PA Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression in
XX a human breast.
XX
XX Claim 25; SEQ ID NO 2967; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
XX The present sequence is one such probe. The probes are useful for
XX measuring human gene expression in a human breast sample, where the probe
XX hybridises at high stringency to a nucleic acid expressed in the human
XX breast. The probes are useful for predicting, diagnosing, grading,
XX staging, monitoring and prognosing diseases of the human breast,
XX particularly those diseases with polygenic aetiology. The diseases
XX include: breast cancer, disorders of development, inflammatory diseases
XX of the breast, fibrocystic changes, proliferative breast disease and non-
XX carcinoma tumours. Note: The sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIFO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 498 BP; 109 A; 140 C; 129 G; 120 T; 0 U; 0 Other;
Query Match 13.0%; Score 314; DB 5; Length 498;
Best Local Similarity 100.0%; Pred. No. 3.2e-143;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

nucleic acid probes for measuring gene expression in a sample derived from human lung comprising single exon nucleic acid probes having one of 12614 nucleic acid sequences mentioned in the specification, or their complements or the 12387 open reading frames derived from the 12614 probes. Also included are a microarray comprising the novel set of probes; the novel set of probes which hybridise at high stringency to a nucleic acid expressed in the human lung; measuring gene expression to a nucleic acid derived from human lung, comprising (a) contacting the array with a collection of detectably labeled nucleic acids derived from human lung mRNA, and (b) measuring the label detectably bound to each probe of the array; identifying exons in a eukaryotic genome, comprising (a) algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagenar syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe of the invention. Note: The sequence data for this patent did not form part of the invention. The specification, but was obtained in electronic format directly from WIPO at fip.wipo.int/pub/published/pct_sequences

Sequence	498 BP;	109 A;	140 C;	129 G;	120 T;	0 U;	0 Other;		
Query Match	13.0%;	Score	314;	DB	6;	Length	498;		
Best Local Similarity	100.0%;	Pred. No.	3.2e-143;						
Matches	314;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1865	AGCTGAAC	TCCAGCGT	GCTGGT	GCCCTAT	GGCGCCGAGT	CTCTCTGAGGACT	CTGCACGAGG	1924
Db	498	AGCTGAACT	CCAGGTGCT	GGTGGCCCT	TATGGCGCCGAGT	CTCTCTGAGGACT	CTGCACGAGG	439	
QY	1925	AGTCAAAGGG	GCTGGGCAAGGAGA	ATGGGA	TGGTAGATT	TGTGAGCT	CCCCACT	CTCCCG	1984
Db	438	AGTCAAAGGG	GCTGGGCAAGGAGA	ATGGGA	TGGTAGATT	TGTGAGCT	CCCCACT	CTCCCG	379
QY	1985	GCCAAGAT	CCGAAGAT	AGGAGGCCA	CTCCGCACGAGCT	TCAAGAAC	CCCATGAC	CCCTAA	2044
Db	378	GCCAAGAT	CCGAAGAT	AGGAGGCCA	CTCCGCACGAGCT	TCAAGAAC	CCCATGAC	CCCTAA	319
QY	2045	ACGGTCT	CTAATAGT	GCAGAC	GACGACAGT	GACCCGAAAGAAAC	CGGCTTAG	CGCCCTGATG	2104
Db	318	ACGGTCT	CTAATAGT	GCAGAC	GACGACAGT	GACCCGAAAGAAAC	CGGCTTAG	CGCCCTGATG	259
QY	2105	GTGCCAGT	GCCTCAAG	CGCAGCTT	GCCTTGCACT	CAGAAAAAT	CCCTTTGCTTAAG	GGCAAAACG	2164
Db	258	GTGCCAGT	GCCTCAAG	CGCAGCTT	GCCTTGCACT	CAGAAAAAT	CCCTTTGCTTAAG	GGCAAAACG	199
QY	2165	GTCTTCT	CTGGAAAG	2178					
Db	198	GTCTTCT	CTGGAAAG	185					

RESULT 28

AAH05659

AAH05659 standard. cDNA: 521 bp

XX

AC	AAH05659;	
XX		
XX	26-JUN-2001 (first entry)	
DE	Human cDNA clone (5'-primer) SEQ ID NO:2494.	
XX		
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss	
XX		
OS	Homo sapiens.	
XX		
PN	EP1074617-A2.	
XX		
PD	07-FEB-2001.	
XX		
PF	28-JUL-2000; 2000EP-00116126.	
XX		
PR	29-JUL-1999; 99JP-00248036.	
PR	27-AUG-1999; 99JP-00300253.	
PR	11-JAN-2000; 2000JP-00118776.	
PR	02-MAY-2000; 2000JP-00183767.	
PR	09-JUN-2000; 2000JP-00241899.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;	
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;	
XX		
DR	WPI; 2001-318749/34.	

SQ Sequence 521 BP; 130 A; 135 C; 122 G; 123 T; 0 U; 11 Other;

RESIST 29

AAH05659

XX

developmental or haematopoietic disorders. The nucleic acid can be used as probes, primers, in biological assays, to determine patterns of gene expression, to design ribozymes and to construct transgenic animals. The present sequence represents one of the 268 disclosed human G-protein-coupled protease cDNA sequences

XX
SQ Sequence 457 BP; 122 A; 113 C; 103 G; 119 T; 0 U; 0 Other;
Query Match 8.5%; Score 205; DB 6; Length 457;
Best Local Similarity 100.0%; Pred. No. 9.4e-90;
Matches 205. Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 138 CTCCGAGGCAGTCTCACTGGAGACATGGATGCAGGTTCTGCCAGCTGGGGGCTGTGTC 197

Qy 198 TTCATTGAATGATGTCTCAAAATCACACACTTTCTTTTAGGACCAGTACTGGTCTGTAGT 257

DB		163	TTCATGGTCTGCGACAGCAGCCGC	
OY		258	TTATTGAGTTTCATCTGTACTGTAATCAAAACCATCACCAAAGGATCAAGCCCT	317

DB
223 TTATTTCAGTTCATCTGTAACCTCAGATAAAATCAAGAACCATCACACATTTAATGCCTGGT
QY
318 AGGTGATGGCATCGCTCTCTCCACAG 342

Db
283 AGGTGATGGCATCGCTCTCTCCACAG 307

RESUL 30
AA115006/C
ID AA115006 standard; DNA; 489 BP.
YY

AC	AA115006;
XX	
DT	12-OCT-2001 (first entry)

DE Probe #4939 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;

RM Cervical Cancer) 55:
 XX
 OS Homo sapiens.
 XX
 XX

PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX

30-JAN-2001; 2001WO-US000670.
PF
XX
04-FEB-2000; 2000US-0180312P.
PR

PR	26-JUN-2000	2000US-00608408.
PR	30-JUN-2000	2000US-00608408.
PR	03-AUG-2000	2000US-00632366.
PR	21-SEP-2000	2000US-0234687P.

PR 27-SEP-2000; 2000QS-0230353F.
PR 04-OCT-2000; 2000GB-00034263.
XX
DA (MOLEC-) MOLECULAR DYNAMICS INC.

XX
PI
XX
Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX

XX Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human cervical epithelial cells.
PT

xx
xx Claim 25; SEQ ID NO 4939; 487pp; English.
ps
pc The present invention relates to human single exon nucleic acid probes
cc

CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC

QY 126 TCAGCCTGGCAGCTCCGAGGCAGTCTCACCTGGAGACATGCATGCAGGTTCTGCCAGCTG 185
|||
Db 121 TCAGCCTGGCAGCTCCGAGGCAGTCTCACCTGGAGACATGCATGCAGGTTCTGCCAGCTG 180
|||

QY 186 GGGGCGCTGTGTCCTTCCATTAATTAACTTTTCTTTAGGACCACTACC 240
Db 181 GGGTGCTGTCTTCATTGAATGATGTGTCAAAACACACACTTTCTTTAGGACCACTACC 240
QY 246 TGGTGCTGTAGTTTTATTTCGAGTTTCATCTGTACCTGATAAATCA 288

```
DB      241 TGGTCCTGAGTAAATCGAGTTCCACCTGCTGTATTACA 289
RESULT 29
```

ABK30444
ID ABK30444 standard; cDNA; 457 BP.
XX
XX

XX
DT 23-APR-2002 (first entry)
XX

DE human G-Protein-Coupled Protein Receptor
XX
KW Human; ss; gene; G-protein-coupled protease; gene therapy; transgenic;
KW protease mediated disorder; proliferative disorder;

KW differential disorder; developmental disorder;
KW haematopoietic disorder.
XX

XX
PN
XX
USC331427-B1.

PD	18-DEC-2001.
XX	
PF	26-MAR-1999; 99US-00280116.
XX	

PR 26-MAR-1999; 99US-00280116.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX

PI Robison KE;
XX
DR WPI; 2002-129545/17.

xx PT New polynucleotides encoding protease homologs of the G-protein-coupled
PT protease family, useful in identifying agonists and antagonists for
PT diagnosis and treatment of protease mediated disorders.
PT

XX Disclosure; Col 243-244; 246pp; English.
PS
XX
XX
XX the invention relates to an isolated human protease nucleic acid molecule

CC comprising a nucleotide sequence of 546 base pairs, one of 268 fully
CC defined in the specification. Also disclosed are production of an
CC isolated polypeptide encoded by the nucleic acid, comprising introducing
CC the nucleic acid into a cell and culturing under conditions to

CC express the protein from the nucleic acid, use of an antibody to detect
CC the encoded protein in a sample and to modulate its *in vivo* activity,
CC identifying agents that bind to the protein and identification of a
CC

CC polynucleotide agent that modulates the expression of a gene product of a gene. The nucleic acid can be used to
CC its complement (i.e. gene therapy). The nucleic acid can be used to
CC identify an agent that modulates the expression or activity of the
CC nucleic acid, and can be used to isolate the protein. The nucleic acid
CC nucleic acid, and can be used to isolate the protein. The nucleic acid

CC can be used in diagnostic assays for determining nucleic acid expression
CC as well as activity in the context of a biological sample (e.g., blood,
CC serum, cells, tissue) to determine whether an individual has a disease or
CC serum, cells, tissue) of developing a disease or disorder associated

CC with aberrant expression or activity of the nucleic acid. The nucleic
CC acid can be used to detect mutations in protease genes and gene
CC expression products such as mRNA. The nucleic acid can be used as
CC a probe for the detection of the expression of the gene in

CC hybridisation proves to be a reasonably good screening method to
CC identify agonists and antagonists that can be used to diagnose and treat
CC such protease mediated disorders e.g., proliferative, differentiative,
CC and other disorders.

CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging of
CC diseases of the cervix, notably cervical cancer. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
Query Match 8.5%; Score 205; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 AGCCCTAGGTGATGGCATGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 371
DB 449 AGCCCTAGGTGATGGCATGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 390
QY 372 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 431
DB 389 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 330
QY 432 TTTTGGCAATGAGCACTGAGTGTAACTTACCTACACACCCTCTTGGCAATACATGCT 491
DB 329 TTTTGGCAATGAGCACTGAGTGTAACTTACCTACACACCCTCTTGGCAATACATGCT 270
QY 492 ATCATGTAACACTCCAAAACATGT 516
DB 269 ATCATGTAACACTCCAAAACATGT 245

RESULT 31
ABA56739/c
ID ABA56739 standard; DNA; 489 BP.
XX AC ABA56739;
XX DT 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #5044.
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US0000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human foetal liver.
XX PS Claim 1; SEQ ID NO 5044; 639pp + Sequence Listing; English.

CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The

CC present sequence is a single exon nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
Query Match 8.5%; Score 205; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 AGCCCTAGGTGATGGCATGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 371
DB 449 AGCCCTAGGTGATGGCATGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTG 390
QY 372 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 431
DB 389 TCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTG 330
QY 432 TTTTGGCAATGAGCACTGAGTGTAACTTACCTACACACCCTCTTGGCAATACATGCT 491
DB 329 TTTTGGCAATGAGCACTGAGTGTAACTTACCTACACACCCTCTTGGCAATACATGCT 270
QY 492 ATCATGTAACACTCCAAAACATGT 516
DB 269 ATCATGTAACACTCCAAAACATGT 245

RESULT 32
AAI36351/c
ID AAI36351 standard; DNA; 489 BP.
XX AC AAI36351;
XX DT 17-OCT-2001 (first entry)
DE Probe #5037 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US0000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX PT gene expression in human placenta.

XX PS Claim 25; SEQ ID NO 5037; 654pp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders

XX SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
Query Match 8.5%; Score 205; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 AGCCCTAGTGTGATGGTCCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTG 371
Db 449 AGCCCTAGTGTGATGGTCCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTG 390
QY 372 TCTTAAGTGGCAACAACACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 431
Db 389 TCTTAAGTGGCAACAACACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 330
QY 432 TTTTGGCAATGACGACTGCAGTGTTTAACTTACACACACCACTCTTCCCAATTTACATGCT 491
Db 329 TTTTGGCAATGACGACTGCAGTGTTTAACTTACACACCACTCTTCCCAATTTACATGCT 270
QY 492 ATCATGTAACACTCCAAAACATGT 516
Db 269 ATCATGTAACACTCCAAAACATGT 245
RESULT 33
ABA46202/c
ID ABA46202 standard; DNA; 489 BP.
XX AC ABA46202;
XX DT 01-FEB-2002 (first entry)
XX DE Human breast cell single exon nucleic acid probe #4897.
XX DE Human; microarray; single exon probe; gene expression; breast; disease;
XX KW cancer; ss.
XX OS Homo sapiens.
XX PN WO200157271-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000662.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX PT New spatially-addressable set of single exon nucleic acid probes, useful
PT for measuring gene expression in sample derived from human breast,
PT comprises number of single exon nucleic acid probes.
XX PS Claim 1; SEQ ID NO 4897; 327pp + Sequence Listing; English.
XX CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting the
CC probes with a collection of detectably labelled nucleic acids derived
CC from mRNA of human breast, and then measuring the label bound to each
CC probe of the microarray. The probes are useful for verifying the
CC expression of regions of genomic DNA predicted to encode proteins. They
CC are useful for gene discovery, and for determining predisposition and/or

CC prognosing breast disease. Gene expression analysis is useful for
CC assessing the toxicity of chemical agents on cells. The microarray of
CC this invention presents a far greater diversity of probes for measuring
CC gene expression, with far less bias than expressed sequence tag
CC microarrays. The method is suitable for rapid production of functional
CC information from genomic sequence. The present sequence is a single exon
CC nucleic acid probe of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
Query Match 8.5%; Score 205; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 312 AGCCCTAGTGTGATGGTCCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTG 371
Db 449 AGCCCTAGTGTGATGGTCCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTG 390
QY 372 TCTTAAGTGGCAACAACACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 431
Db 389 TCTTAAGTGGCAACAACACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 330
QY 432 TTTTGGCAATGACGACTGCAGTGTTTAACTTACACACCACTCTTCCCAATTTACATGCT 491
Db 329 TTTTGGCAATGACGACTGCAGTGTTTAACTTACACACCACTCTTCCCAATTTACATGCT 270
QY 492 ATCATGTAACACTCCAAAACATGT 516
Db 269 ATCATGTAACACTCCAAAACATGT 245
RESULT 34
AAK30399/c
ID AAK30399 standard; DNA; 489 BP.
XX AC AAK30399;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed single exon probe SEQ ID NO: 4956.
XX DE Human; bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-489900/53.
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 4956; 658pp + Sequence Listing; English.
PS

XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
 CC the probes of the invention
 XX
 SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;

Query Match 8.5%; Score 205; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 9.3e-90;
 Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 AGCCCTAGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTCGAGAGATTG 371
 DB |||||
 QY 449 AGCCCTAGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTCGAGAGATTG 390
 DB |||||
 QY 372 TCTTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 431
 DB |||||
 QY 389 TCTTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 330
 DB |||||
 QY 432 TTTTGCCTAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 491
 DB |||||
 QY 329 TTTTGCCTAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 270
 DB |||||
 QY 492 ATCATGTAACACTCCAAACATGT 516
 DB |||||
 QY 269 ATCATGTAACACTCCAAACATGT 245

RESULT 35
 ABS30038/c
 ID ABS30038 standard; DNA; 489 BP.
 XX
 AC ABS30038;

25-FEB-2003 (first entry)
 Human liver single exon probe, SEQ ID NO 5028.
 Human; single exon nucleic acid probe; liver; cirrhosis;
 hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
 coronary heart disease; ss.

Homo sapiens.
 WO200157273-A2.
 09-AUG-2001.
 30-JAN-2001; 2001WO-US000664.
 04-FEB-2000; 2000US-0180312P.
 26-MAY-2000; 2000US-0207456P.
 30-JUN-2000; 2000US-00608408.
 03-AUG-2000; 2000US-00632366.
 21-SEP-2000; 2000US-0234687P.
 27-SEP-2000; 2000US-0236359P.
 04-OCT-2000; 2000GB-00024263.
 (MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-488898/53.

Human genome-derived single exon nucleic acid probes useful for analyzing
 gene expression in human adult liver.
 Claim 1; SEQ ID NO 5028; 658pp; English.
 The invention relates to a single exon nucleic acid probe (SEN) (I) for

CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 13109 defined nucleotide sequences given in the
 CC specification (or complements/fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult liver.
 CC (I) may be used for predicting, measuring and displaying gene expression
 CC in samples derived from human adult liver. The genes identified may be
 CC involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
 CC associated with coronary heart disease. ABS25011-ABS51005 represent human
 CC liver single exon nucleic acid probes of the invention. Note: The
 CC sequence information for this patent does not appear in the printed
 CC specification but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 489 BP; 162 A; 97 C; 94 G; 136 T; 0 U; 0 Other;
 Query Match 8.5%; Score 205; DB 4; Length 489;
 Best Local Similarity 100.0%; Pred. No. 9.3e-90;
 Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 AGCCCTAGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTCGAGAGATTG 371
 DB |||||
 QY 449 AGCCCTAGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTCGAGAGATTG 390
 DB |||||
 QY 372 TCTTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 431
 DB |||||
 QY 389 TCTTAAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 330
 DB |||||
 QY 432 TTTTGCCTAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 491
 DB |||||
 QY 329 TTTTGCCTAGTGGCAACAACCTCATAGAGTTGGAGCTGGCTCCAGAAATTTGGCAATACCTG 270
 DB |||||
 QY 492 ATCATGTAACACTCCAAACATGT 516
 DB |||||
 QY 269 ATCATGTAACACTCCAAACATGT 245

RESULT 36
 AA104774/c
 ID AA104774 standard; DNA; 489 BP.
 XX
 AC AA104774;
 DT 09-OCT-2001 (first entry)
 DE
 DE Probe #4765 used to measure gene expression in human breast sample.
 KW Probe: human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 XX
 OS Homo sapiens.
 XX
 XX WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 29-JAN-2001; 2001WO-US000661.
 XX
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.
 Penn SG, Hanzel DK, Chen W, Rank DR;
 WPI; 2001-476286/51.
 Novel single exon nucleic acid probe used to measuring gene expression in

Db 269 ATCATGAACTCCAAACATGT 245

RESULT 38
AAI24193/c
ID AAI24193 standard; DNA; 198 BP.
XX AC AAI24193;
XX DT 12-OCT-2001 (first entry)
DE Probe #14126 for gene expression analysis in human cervical cell sample.
XX KW Probe; human; microarray; gene expression; cervical epithelial cell;
XX KW cervical cancer; ss.
XX OS Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000670.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human cervical epithelial cells.
XX Claim 25; SEQ ID NO 14126; 487pp; English.
XX The present invention relates to human single exon nucleic acid probes
XX (SENPs). The present sequence is one such probe. The SENPs are derived
XX from human HeLa cells. The SENPs can be used to produce a single exon
XX microarray, which can be used for measuring human gene expression in a
XX sample derived from human cervical epithelial cells. By measuring gene
XX expression, the probes are therefore useful in grading and/or staging of
XX diseases of the cervix, notably cervical cancer. Note: The sequence data
XX for this patent did not form part of the printed specification, but was
XX obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
Query Match 8.2%; Score 198; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 CCTAGGTGATGGCATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 373
Db 198 CCTAGGTGATGGCATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 139
QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 79
QY 434 TTGCCAATGAGCACTGAGTGTGTTAACTTACCTACACACCACCTCTTGCCAATTATGCTAT 493
Db 78 TTGCCAATGAGCACTGAGTGTGTTAACTTACCTACACACCACCTCTTGCCAATTATGCTAT 19
QY 494 CACATGAACACTCCAAAA 511

Db 18 CACATGAACACTCCAAAA 1

RESULT 39
ABA69312/c
ID ABA69312 standard; DNA; 198 BP.
XX AC ABA69312;
XX DT 01-FEB-2002 (first entry)
DE Human foetal liver single exon nucleic acid probe #17617.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 4; SEQ ID NO 17617; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a single exon nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;
Query Match 8.2%; Score 198; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 314 CCTAGGTGATGGCATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 373
Db 198 CCTAGGTGATGGCATCGCTCCCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 139
QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAAATTGGGCAATACCTGTT 79
QY 434 TTGCCAATGAGCACTGAGTGTGTTAACTTACCTACACACCACCTCTTGCCAATTATGCTAT 493
Db 78 TTGCCAATGAGCACTGAGTGTGTTAACTTACCTACACACCACCTCTTGCCAATTATGCTAT 19
QY 494 CACATGAACACTCCAAAA 511
Db 18 CACATGAACACTCCAAAA 1

ABAS1313 standard; DNA; 198 BP.

ABAS1313;

01-FEB-2002 (first entry)

Human breast cell single exon nucleic acid probe #10008.

Human; microarray; single exon probe; gene expression; breast; disease; cancer; ss.

Homo sapiens.

WO200157271-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US003662.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-496933/54.

New spatially-addressable set of single exon nucleic acid probes, useful for measuring gene expression in sample derived from human breast, comprises number of single exon nucleic acid probes.

Claim 4; SEQ ID NO 10008; 327pp + Sequence Listing; English.

The invention relates to a spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human breast and Br 474 cells. The method involves contacting the probes with a collection of detectably labelled nucleic acids derived from mRNA of human breast, and then measuring the label bound to each probe of the microarray. The probes are useful for verifying the expression of regions of genomic DNA predicted to encode proteins. They are useful for gene discovery, and for determining predisposition and/or prognosing breast disease. Gene expression analysis is useful for assessing the toxicity of chemical agents on cells. The microarray of this invention presents a far greater diversity of probes for measuring gene expression, with far less bias than expressed sequence tag microarrays. The method is suitable for rapid production of functional information from genomic sequence. The present sequence is a single exon nucleic acid probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match 8.2%; Score 198; DB 4; Length 198;

Best Local Similarity 100.0%; Pred. No. 2.6e-86;

Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 373

Db 198 CCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 139

QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTT 433

Db 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTT 79

QY 434 TTGCCAATGCAGCTGAGTGTAACTACACACCACCTCTTCCCAATTACATGCTAT 493

RESULT 40

AAI49479/c

ID AAI49479 standard; DNA; 198 BP.

AC AAI49479;

17-OCT-2001 (first entry)

Probe #18165 used to measure gene expression in human placenta sample.

Probe; microarray; human; placenta; antenatal diagnosis; genetic disorder; ss.

Homo sapiens.

WO200157272-A2.

09-AUG-2001.

30-JAN-2001; 2001WO-US000663.

04-FEB-2000; 2000US-0180312P.

26-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-488897/53.

Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human placenta.

Claim 25; SEQ ID NO 18165; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENP). The present sequence is one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders

Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match 8.2%; Score 198; DB 4; Length 198;

Best Local Similarity 100.0%; Pred. No. 2.6e-86;

Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 373

Db 198 CCTAGGTGATGGCATCGCTCTCCACAGAAAGTTCTTTCCCATCTGAGAGATTGTC 139

QY 374 TTAAGTGGCAACAACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTT 433

Db 138 TTAAGTGGCAACAACTCATAGATTGGAGCTGGCTCCAGAAATTTGGCAATACCTGTT 79

QY 434 TTGCCAATGCAGCTGAGTGTAACTACACACCACCTCTTCCCAATTACATGCTAT 493

Db 78 TTGCCAATGCAGCTGAGTGTAACTACACACCACCTCTTCCCAATTACATGCTAT 19

QY 494 CACATGAACACTCCAAA 511

Db 18 CACATGAACACTCCAAA 1

RESULT 41

ABAS1313/c

```
Db      78  TTGCAATGCAGCTGCAGTGTTTAACTTACATACACACCTCTTGCCCAATTACATGCTAT 19
Qy      494 CACATGAACACTCCAAAA 511
Db      18  CACATGAACACTCCAAAA 1

RESULT 42
AAK43414/c
ID      AAK43414 standard; DNA; 198 BP.
XX
AC      AAK43414;
XX
DE      06-NOV-2001 (first entry)
XX
DE      Human bone marrow expressed single exon probe SEQ ID NO: 17971.
XX
KW      Human; bone marrow expressed exon; gene expression analysis; probe;
KW      microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157276-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000668.
XX
PR      04-FEB-2000; 2000US-0180312P.
PR      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488900/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for analyzing
PT      gene expression in human bone marrow.
XX
PS      Example 4; SEQ ID NO 17971; 658pp + Sequence Listing; English.
XX
CC      The present invention provides a number of single exon nucleic acid
CC      probes which are derived from genomic sequences expressed in the human
CC      bone marrow. They can be used to measure gene expression in bone marrow
CC      samples, which may enable the improved diagnosis and treatment of cancers
CC      such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC      the probes of the invention
XX
SQ      Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match      8.2%; Score 198; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      314  CCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTGTC 373
Db      198  CCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTGTC 139

Qy      374  TTAAGTGGCAAACTCATAGAGTTGGAGTGGGCTCCAGAAATTTGGCAATACCTGTT 433
Db      138  TTAAGTGGCAAACTCATAGAGTTGGAGTGGGCTCCAGAAATTTGGCAATACCTGTT 79

Qy      434  TTGCAATGCAGACTGCAGTGTTTAACTTACATACACACCTCTTGCCCAATTACATGCTAT 493
Db      78  TTGCAATGCAGACTGCAGTGTTTAACTTACATACACACCTCTTGCCCAATTACATGCTAT 19
```

```
Qy      494 CACATGAACACTCCAAAA 511
Db      18  CACATGAACACTCCAAAA 1

RESULT 43
ABS43035/c
ID      ABS43035 standard; DNA; 198 BP.
XX
AC      ABS43035;
XX
DT      25-FEB-2003 (first entry)
XX
DE      Human liver single exon probe, SEQ ID No 18025.
XX
KW      Human; single exon nucleic acid probe; liver; cirrhosis;
KW      hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW      coronary heart disease; ss.
XX
OS      Homo sapiens.
XX
PN      WO200157273-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US000664.
XX
PR      04-FEB-2000; 2000US-0180312P.
PR      26-MAY-2000; 2000US-0207456P.
PR      30-JUN-2000; 2000US-00608408.
PR      03-AUG-2000; 2000US-00632366.
PR      21-SEP-2000; 2000US-0234687P.
PR      27-SEP-2000; 2000US-0236359P.
PR      04-OCT-2000; 2000GB-00024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR      WPI; 2001-488998/53.
XX
PT      Human genome-derived single exon nucleic acid probes useful for analyzing
PT      gene expression in human adult liver.
XX
PS      Claim 4; SEQ ID NO 18025; 658pp; English.
XX
CC      The invention relates to a single exon nucleic acid probe (SENp) (I) for
CC      measuring human gene expression in a sample derived from human adult
CC      liver, comprising one of 13109 defined nucleotide sequences given in the
CC      specification (or complements/ fragments). The probe hybridises at high
CC      stringency to a nucleic acid molecule expressed in the human adult liver.
CC      (I) may be used for predicting, measuring and displaying gene expression
CC      in samples derived from human adult liver. The genes identified may be
CC      involved in genetic liver diseases such as cirrhosis,
CC      hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC      associated with coronary heart disease. ABS25011-ABS51005 represent human
CC      liver single exon nucleic acid probes of the invention. Note: The
CC      sequence information for this patent does not appear in the printed
CC      specification but was obtained in electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match      8.2%; Score 198; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      314  CCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTGTC 373
Db      198  CCTAGGTGATGGCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTGTC 139

Qy      374  TTAAGTGGCAAACTCATAGAGTTGGAGTGGGCTCCAGAAATTTGGCAATACCTGTT 433
```

Db 138 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 79
QY 434 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTGGCAATTAATCATGCTAT 493
Db 78 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTGGCAATTAATCATGCTAT 19
QY 494 CACATGAACACTCCAAA 511
Db 18 CACATGAACACTCCAAA 1

RESULT 44
AAI09758/c
ID AAI09758 standard; DNA; 198 BP.
XX
AC AAI09758;
XX
DT 09-OCT-2001 (first entry)
XX
DE Probe #9749 used to measure gene expression in human breast sample.
XX
KW Probe; human; breast disease; breast cancer; development disorder; ss;
KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
XX
OS Homo sapiens.
XX
PN WO200157270-A2.
XX
PD 09-AUG-2001.
XX
PF 29-JAN-2001; 2001WO-US000661.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2001-476286/51.
XX
PT Novel single exon nucleic acid probe used to measuring gene expression in
PT a human breast.
XX
PS Claim 25; SEQ ID NO 9749; 322pp; English.
XX
CC The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridises at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and non-
CC carcinoma tumours. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match 8.2%; Score 198; DB 5; Length 198;
Best Local Similarity 100.0%; Fred. No. 2.6e-86;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CCTAGGTGATGCGATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 373
Db 198 CCTAGGTGATGCGATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAGATTGTC 139

QY 374 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGCAATACCTGTT 79
QY 434 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTGGCAATTAATCATGCTAT 493
Db 78 TTGCAATGCGACACTGCGAGTGTTTAACTACACACCACTCTTGGCAATTAATCATGCTAT 19
QY 494 CACATGAACACTCCAAA 511
Db 18 CACATGAACACTCCAAA 1

RESULT 45
ABS17512/c
ID ABS17512 standard; DNA; 198 BP.
XX
AC ABS17512;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID No 17503.
XX
KW Human; ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US000665.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-00608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to
PT measure gene expression in human lung samples.
XX
PS Claim 4; SEQ ID NO 17503; 634pp; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of probes
CC; the novel set of probes which hybridise at high stringency to a nucleic
CC acid expressed in the human lung; measuring gene expression in a sample
CC derived from human lung, comprising (a) contacting the array with a
CC collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of the
CC array; identifying exons in a eukaryotic genome, comprising (a)

algorithmically predicting at least one exon from genomic sequences of the eukaryote; and (b) detecting specific hybridisation of detectably labeled nucleic acids from eukaryote lung mRNA, to a single exon probe, having a fragment identical to the predicted exon, the probe is included in the above mentioned microarray; assigning exons to a single gene, comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagenier syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a single exon probe open reading frame of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 198 BP; 53 A; 37 C; 53 G; 55 T; 0 U; 0 Other;

Query Match
Best Local Similarity 8.2%; Score 198; DB 6; Length 198;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CCTAGGTCATGGCATCGTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTC 373
Db |||||
QY 198 CCTAGGTCATGGCATCGTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTC 139
Db |||||
QY 374 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTGGCAATACCTGTT 433
Db |||||
QY 138 TTAAGTGGCAACAACTCATAGAGTTGGAGCTGGCTCCAGAAATTGGCAATACCTGTT 79
Db |||||
QY 434 TTGCCAATGCAGCACTGCAGTGTAACTACACACACCCTCTTGCCAAATTACATGCTAT 493
Db |||||
QY 78 TTGCCAATGCAGCACTGCAGTGTAACTACACACACCCTCTTGCCAAATTACATGCTAT 19
Db |||||
QY 494 CACATGAACACTCCAAA 511
Db |||||
QY 18 CACATGAACACTCCAAA 1

RESULT 46
ABQ95217
ID ABQ95217 standard; DNA; 213 BP.
AC ABQ95217;
DT 28-OCT-2002 (first entry)
DE Tumour suppression-related oligonucleotide #868.
KW Tumour; cytostatic; antiviral; neuroprotective; nootropic; neuroleptic;
KW tumour suppression; tumour reversion; apoptosis; viral resistance; human;
KW viral infection; cell degeneration disease; neurodegeneration; ds;
KW Alzheimer's disease; schizophrenia; immune disease; inflammatory disease.
OS Homo sapiens.
PN FR2819824-A1.
XX
XX
PD 26-JUL-2002.
XX
XX 23-JAN-2001; 2001FR-00000899.

PR 23-JAN-2001; 2001FR-00000899.
XX
PA (MOLE-) MOLECULAR ENGINES LAB SA.
XX
PI Teletman A, Anson R, Tuijnder M, Susini L;
XX WPI; 2002-610803/66.
XX
PT New nucleic acid implicated e.g. in tumor suppression, useful for
PT diagnosis of tumors, viral infection and cellular degeneration and for
PT drug screening.
XX
PS Claim 1; Page 256; 623pp; French.
XX
CC The present invention relates to novel human nucleic acid sequences (I).
CC The present sequence is one such nucleic acid sequence. Expression of (I)
CC are implicated in tumour suppression or reversion and apoptosis and viral
CC resistance. (I) are useful as probes or primers for detecting,
CC identifying, measuring and/or amplifying nucleic acid sequences, as
CC antisense reagents and for recombinant production of polypeptides. (I),
CC polypeptides (II) encoded by (I), vector containing (I), cells containing
CC these vectors and antibodies (Ab) against (II) are all useful for
CC treatment/prevention of viral, tumour and cell degeneration diseases
CC (especially neurodegeneration, such as Alzheimer's disease and
CC schizophrenia). Analysing the expression of (I) is also useful for
CC diagnosis and/or prognosis of such diseases. Transgenic animals carrying
CC (I) are used for studying the aetiology of these diseases (also immune
CC and inflammatory diseases). Note: In the present specification, SEQ ID 1
CC to 2280 are claimed in Claim 1, however only SEQ ID 1 to 2270 are shown
CC in the specification
XX
SQ Sequence 213 BP; 64 A; 44 C; 47 G; 58 T; 0 U; 0 Other;

Query Match
Best Local Similarity 5.7%; Score 137; DB 6; Length 213;
Matches 137; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 GCTTGATGGAGAAAACCTCGTACAGTCGAGCAAGTGTAAAAGATGGTTCCAGTTCAA 977
Db |||||
QY 70 GCTTGATGGAGAAAACCTCGTACAGTCGAGCAAGTGTAAAAGATGGTTCCAGTTCAA 129
Db |||||
QY 978 GAGGTTCACTATCCATAGATCTCTTAATGTTTACACTTCTCTGAAACGTTTGCAAA 1037
Db |||||
QY 130 GAGGTTCACTATCCATAGATCTCTTAATGTTTACACTTCTCTGAAACGTTTGCAAA 1037
Db |||||
QY 1038 TTTTACCGGTGAAAAA 1054
Db |||||
QY 190 TTTTACCGGTGAAAAA 206
Db |||||

RESULT 47
ABN62290/c
ID ABN62290 standard; cDNA; 292 BP.
XX
XX
AC ABN62290;
XX
XX
DT 28-JUN-2002 (first entry)
DE Human cancer related polynucleotide SEQ ID NO 2257.
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200214500-A2.
XX
XX 21-FEB-2002.
XX
XX 16-AUG-2001; 2001WO-US025840.
XX
XX 16-AUG-2000; 2000US-0226326P.

PA	(CHIR) CHIRON CORP.
PPA	(HYSE-) HYSEQ INC.
XX	
PI	Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PPI	Lamson G, Scott EM, Zhang G, Kassam A, Pot D, Labat I;
PI	
XX	WPI; 2002-241905/29.
DR	
XX	New nucleic acid for producing a polypeptide, detecting differentially
PT	expressed genes correlated with a cancerous state of a mammalian cell,
PT	and inhibiting tumor growth.
PT	
XX	
PS	Claim 1; SEQ ID NO 2257; 883pp + Sequence Listing; English.
CC	
XX	The invention relates to an isolated polynucleotide (ABN27253-ABN3262)
CC	with cytosatic activity. The polynucleotide is used to produce a
CC	polypeptide, to detect differentially expressed genes correlated with a
CC	cancerous state of a mammalian cell and to inhibit tumour growth. The
CC	polynucleotide is used as a probe in mapping and tissue profiling. The
CC	encoded polypeptide and antibodies to the polypeptide can also be used
CC	for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC	gene therapy. Note: The sequence data for this patent did not form part
CC	of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
SQ	Sequence 292 BP; 103 A; 65 C; 62 G; 62 T; 0 U; 0 Other;
	Query Match 4.1%; Score 100; DB 6; Length 292;
	Best Local Similarity 99.3%; Pred. No. 3.1e-38;
	Matches 150; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy	312 AGCCTAGGTGATGGCATCGCTCTCCACAGAAAAGTTCTTTTCCCATCTCAGAAGATTG 371
Dd	151 AGCCTAGGTGATGGCATCGCTCTCCACAGAAAAGTTCTTTTCCCATCTCAGAAGATTG 92
Qy	372 TCCTTAAGTGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAAATTGGCAATACCTG 431
Dd	91 TCCTTAAGTGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAAATTGGCAATACCTG 32
Qy	432 TTTTGCCAATGCAGCACTGCAGTGTTTAAC 462
Dd	31 TTTTGCCAATGCAGCACTGCAGTGTTTAAC 1
RESULT 48	
AAK57307	
ID	AAK57307 standard; cDNA; 657 BP.
XX	
AC	AAK57307;
XX	
DT	
XX	06-NOV-2001 (first entry)
XX	Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:2367.
DE	
XX	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW	Cytostatic; gene therapy; vaccine; metastasis; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200157182-A2.
XX	
PD	09-AUG-2001.
XX	
XX	17-JAN-2001; 2001WO-US0001354.
XX	
XX	31-JAN-2000; 2000US-0179065P.
PR	04-FEB-2000; 2000US-0180628P.
PR	24-FEB-2000; 2000US-0184664P.
PR	02-MAR-2000; 2000US-0186350P.
PR	16-MAR-2000; 2000US-0189874P.
PR	17-MAR-2000; 2000US-0190076P.
PR	18-APR-2000; 2000US-0198123P.
PR	19-MAY-2000; 2000US-0205515P.

Wed Aug 18 13:52:58 2004

```

XX  Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J;
PI  Hillman JL, Baughn MR, Yao MG, Burford N, Batra S, Policky JG;
XX
XX  WPI; 2002-034502/04.
DR  P-PSDB; AAE15241.
XX
XX  New human RNA metabolism protein for diagnosing or treating nervous
PT  system disorders, autoimmune/inflammatory disorders, cell proliferative
PT  disorders and developmental disorders.
XX
XX  Claim 11; Page 172; 196pp; English.
XX
XX  The invention relates to human RNA metabolism proteins (RMEP) and their
CC  corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,
CC  treating and preventing nervous system disorders (epilepsy, dementia,
CC  stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);
CC  prion diseases; fatal familial insomnia, nutritional and metabolic
CC  diseases of the nervous system; inherited, metabolic, endocrine and toxic
CC  myopathy; mental disorders (mood, anxiety, schizophrenic disorders)
CC  amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-
CC  acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,
CC  atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,
CC  Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
CC  arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,
CC  ulcerative colitis, and infections); cell proliferative disorders (cancer
CC  arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental
CC  disorders (renal tubular acidosis). RMEP DNA is useful in drug screening
CC  techniques, gene therapy and for creating transgenic animals. The present
CC  sequence is human RMEP-4 cDNA
XX
XX  Sequence 637 BP; 175 A; 130 C; 150 G; 182 T; 0 U; 0 Other;
SQ
Query Match 1.2%; Score 30; DB 6; Length 637;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
DB 37 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 8
RESULT 50
AAD08062
XX  AAD08062 standard; cDNA; 733 BP.
XX
XX  06-AUG-2001 (first entry)
XX
XX  Human extracellular matrix and cell adhesion molecule-18 (XMAD-18) cDNA.
XX
XX  Human; extracellular matrix and cell adhesion molecule; XMAD;
XX  Gene therapy; genetic disorder; adrenoleukodystrophy; leukaemia;
XX  Down's syndrome; cystic fibrosis; Gaucher's disease; myotonic dystrophy;
XX  sickle cell anaemia; thalassaemia; autoimmune disorder; adenocarcinoma;
XX  inflammatory disorder; acquired immune deficiency syndrome; AIDS;
XX  Addison's disease; allergy; anaemia; asthma; atherosclerosis; melanoma;
XX  Crohn's disease; diabetes mellitus; atopic dermatitis; lymphoma; cancer;
XX  glomerulonephritis; multiple sclerosis; Grave's disease; osteoarthritis;
XX  osteoporosis; psoriasis; rheumatoid arthritis; ulcerative colitis;
XX  infection; cell proliferative disorder; actinic keratosis; myeloma;
XX  arteriosclerosis; neurotropic; anticonvulsant; antithyroid; nephrotropic;
XX  neuroprotective; dermatological; chromosome 22; ss.
XX
XX  Homo sapiens.
XX
XX  Location/Qualifiers
XX  Key 146..703
XX  CDS /*tag= a
XX  /product= "Human extracellular matrix and cell adhesion
XX  molecule (XMAD)"

```

```

FN WO200142285-A2.
XX
XX  14-JUN-2001.
XX
XX  05-DEC-2000; 2000WO-US032990.
XX
XX  10-DEC-1999; 99US-0172852P.
XX  16-DEC-1999; 99US-0172354P.
XX  (INCY-) INCYTE GENOMICS INC.
XX
XX  Yue H, Tang YT, Lal P, Burford N, Azimzai Y, Patterson C;
PI  Baughn MR, Lu DAM, Shah P, Au-Young J;
XX
XX  WPI: 2001-381632/40.
XX  P-PSDB; AAE03854.
XX
XX  New human extracellular matrix and cell adhesion molecules and
PT  polynucleotide sequences encoding them, useful for diagnosis, prevention,
PT  treatment of genetic, autoimmune and cell proliferative disorders.
XX
XX  Claim 5; Page 131; 135pp; English.
XX
XX  The present cDNA sequence encodes human extracellular matrix and cell
CC  adhesion molecule (XMAD). The XMAD is used for screening a compound for
CC  effectiveness as an agonist or antagonist of XMAD. The identified agonist
CC  or antagonist are used for treating a disease or condition associated
CC  with decreased or increased expression of functional XMAD. The
CC  polynucleotides encoding XMAD are useful in somatic or germline gene
CC  therapy to correct a genetic deficiency, to express a conditionally
CC  lethal gene product and to express a protein which affords protection
CC  against intracellular parasites and also for diagnosis of disorders
CC  associated with expression of XMAD. They are also used for generating
CC  hybridisation probes useful in mapping the naturally occurring genomic
CC  sequences and to create knock in humanised animals (pigs) or transgenic
CC  animals (mice or rats) to model human diseases. Oligonucleotide or longer
CC  fragments derived from the polynucleotide sequences may be used as
CC  elements on a microarray. Antibodies which specifically bind XMAD may be
CC  used for the diagnosis of disorders associated with the expression of
CC  XMAD, or in assays to monitor patients being treated with XMAD. Diseases
CC  diagnosed, prevented or treated include genetic disorders such as
CC  adrenoleukodystrophy, Down's syndrome, cystic fibrosis, Gaucher's
CC  disease, myotonic dystrophy, sickle cell anaemia, thalassaemia,
CC  autoimmune/inflammatory disorders such as acquired immune deficiency
CC  syndrome (AIDS), Addison's disease, diabetes mellitus, atopic dermatitis,
CC  atherosclerosis, Crohn's disease, diabetes mellitus, ulcerative colitis,
CC  glomerulonephritis, multiple sclerosis, rheumatoid arthritis, infectious
CC  osteoporosis, psoriasis, parasitic, protozoal and helminthic infections and
CC  bacterial, fungal, proliferative disorders such as actinic keratosis, arteriosclerosis
CC  cell proliferative disorders such as bladder, bone marrow, brain and uterus
CC  and cancer including breast, lymphoma, melanoma and myeloma
XX
XX  Sequence 733 BP; 107 A; 240 C; 245 G; 141 T; 0 U; 0 Other;
SQ
Query Match 1.2%; Score 30; DB 4; Length 733;
Best Local Similarity 100.0%; Pred. No. 0.00065;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
DB 696 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 725
Search completed: August 17, 2004, 00:36:39
Job time : 1367 secs

```


Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 4596.f. For more information about this cluster, see <http://www.genoscope.cns.fr/>
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DF001AD12QP2&cluster=4596.f>. Contact : Feng Liang Email : liang@lifetech.com URL : <http://fulllength.invitrogen.com/> Invitrogen Corporation 1600 <http://www.fulllength.invitrogen.com/> ID : CS0DF001AD12QP2.

FEATURES
source

```

1. 1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODF001YG23"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/notes="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo (gr) primer. Five prime end
was enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

```

ORIGIN

```
Query Match      37.7%; Score 909; DB 13; Length 1201;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 959; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

23	CGCGCGCCGAGGGGGATGGAGCGAGCGCGGAGCGGGTCAAGTTGAACAATGACCAATAG	82
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36	CGCGCGCCGAGGGGGATGGAGCGAGCGCGGAGCGGGTCAAGTTGAACAATGACCAATAG	125
37		
38		
39		
40		
41		
42		
43	TTGACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGAAATCAGCTGGCGAGCTCCG	142
44		
45		
46		
47		
48		
49		
50		
51		
52		
53	AGCGAGTCTCACTGAGAGCATGGATGACAGTTCTGCCAGCTGGGTGCTGTGCTTCAT	202
54		
55		
56		
57		
58	186 AGCGAGTCTCACTGAGAGCATGGATGACAGTTCTGCCAGCTGGGTGCTGTGCTTCAT	245
59		
60		
61		
62		
63	203 TGAATCATGTGTCAATATCACACACTTCTTTAGGACCACTACCTGGTGTCTAGTCTTATT	262
64		
65		
66	246 TGAATGATGTGTCAATATCACACACTTCTTTAGGACCACTACCTGGTGTCTAGTCTTATT	305
67		
68		
69		
70		
71		
72	263 CGAGTTTCATCTGTACCTGATATAATCAAAACCATCAACAAAGGATCAAGCCCTAGGTG	322
73		
74		
75	306 CGAGTTTCATCTGTACCTGATATAATCAAAACCATCAACAAAGGATCAAGCCCTAGGTG	365
76		
77		
78	323 ATGGCATTCGCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAGAGATTTGTCTTAAGTGGC	382
79		
80		
81	366 ATGGCATTCGCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAGAGATTTGTCTTAAGTGGC	425
82		
83		
84		
85	383 AACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCCAATG	442
86		
87		
88	426 AACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTGGCCAATG	485
89		
90		
91	443 CAGCACTGCAAGTGTTTAACTACACACCACCTCTTGCCAAATTACAGTCTACATGAAC	502
92		
93	486 CAGCACTGCAAGTGTTTAACTACACACCACCTCTTGCCAAATTACAGTCTACATGAAC	545
94		
95	503 ACTCCAAAAACATGTAATGCAGAGGGCTTTTGATGATGTGTAACATGTCGAAGCACATATTA	562
96	546 ACTCCAAAAACATGTAATGCAGAGGGCTTTTGATGATGTGTAACATGTCGAAGCACATATTA	605
97		
98	563 CCGAGGCACTCAGTAATCTCGGGGAGTATTATAACCAATGTTGTGATCAATGAGATGC	622
99	606 CCGAGGCACTCAGTAATCTCGGGGAGTATTATAACCAATGTTGTGATCAATGAGATGC	665
100		
101	623 GCGGTATAGCTAGGCACCTCCGTTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAAT	682
102		
103	666 GCGGTATAGCTAGGCACCTCCGTTTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAAT	725
104		
105	683 ACACATGTTGATGCTATGCAGAAAGCATGCTTGTGAATGGCAGCAATAAATTAGACAGACACA	742

C	78	96	4.0	602	14	CB217660	CB217660 NISC_rb04
C	79	91	3.8	975	13	AA248763	AA248763 j4452.seq
C	80	91	3.8	941	13	BX366515	BX366515 BX366515
C	81	88	3.6	408	28	AQ181035	AQ181035 HS_3219_B
C	82	66	2.8	711	14	CF178635	CF178635 808125_MA
C	83	66	2.7	75	14	CA513635	CA513635 10-3_1A1
C	84	66	2.7	75	14	CA513637	CA513637 10-3_1A6
C	85	64	2.7	453	28	AQ825501	AQ825501 HS_5499_A
C	86	60	2.5	468	28	AQ603902	AQ603902 HS_2135_A
C	87	59	2.4	705	14	CF178303	CF178303 807357_MA
C	88	56	2.3	667	14	CB422965	CB422965 596130_MA
C	89	53	2.3	687	14	CB428011	CB428011 603659_MA
C	90	53	2.2	415	10	BE079835	BE079835 RC6_BT062
C	91	47	1.9	409	10	BF561189	BF561189 UI-R-C0-h
C	92	47	1.9	550	29	CF484116	CF484116 tigr-gss-
C	93	46	1.9	537	10	BF523269	BF523269 UI-R-C3-t
C	94	46	1.9	685	10	BE651621	BE651621 UI-M-BH2
C	95	41	1.7	502	10	BF461120	BF461120 UI-M-CG0P
C	96	41	1.7	532	10	BF404275	BF404275 UI-R-CAL-
C	97	41	1.7	588	10	BF079818	BF079818 H3045D11-
C	98	41	1.7	599	13	C79892	C79892 C79892_Mous
C	99	41	1.7	620	12	BM940384	BM940384 UI-M-CG0P
C	100	41	1.7	644	13	BUE10706	BUE10706 UI-M-FCO-
C	101	41	1.7	672	14	CD350154	CD350154 UI-M-FYO-
C	102	41	1.7	703	13	BQ770634	BQ770634 UI-M-FIO-
C	103	41	1.7	729	14	CF532651	CF532651 UI-M-GHO-
C	104	41	1.7	735	14	CB521515	CB521515 UI-M-GHO-
C	105	41	1.7	771	10	BG066566	BG066566 H3045D11-
C	106	41	1.7	773	14	CB951919	CB951919 UI-M-GHO-
C	107	41	1.7	799	14	CF535655	CF535655 UI-M-GHO-
C	108	41	1.7	1485	11	AKO10801	AKO10801 Mus_muscu
C	109	40	1.7	662	29	AGI16029	AGI16029 Pan_trogl
C	110	38	1.6	318	13	BY129297	BY129297 BY129297
C	111	38	1.6	347	13	BY010203	BY010203 BY010203
C	112	38	1.6	348	13	BY010350	BY010350 BY010350
C	113	38	1.6	391	13	BY050048	BY050048 BY050048
C	114	38	1.6	397	13	BY300193	BY300193 BY300193
C	115	38	1.6	402	13	BY013984	BY013984 BY013984
C	116	38	1.6	412	13	BY018798	BY018798 BY018798
C	117	38	1.5	668	13	BY734416	BY734416 BY734416
C	118	37	1.5	417	10	AW514181	AW514181 h24g06.x
C	119	37	1.5	446	10	BF400191	BF400191 UI-R-CAL-
C	120	34	1.4	750	12	BI108038	BI108038 602903424
C	121	32	1.3	618	29	CE119459	CE119459 tigr-gss-
C	122	32	1.3	724	28	BZ247062	BZ247062 CH230-419
C	123	31	1.3	438	14	CB790275	CB790275 AMGNUTC:S
C	124	30	1.2	1977	14	CD013953	CD013953 90137783
C	125	30	1.2	1987	10	AW888223	AW888223 MXRA8_Hum

ATTACHMENTS

RESULT 1
BX418055
LOCUS
DEFINITION
BX418055 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
CDSD001YG23 5-PRIME, mRNA sequence.
ACCESSION
BX418055
VERSION
BX418055.1 GI:30646403
EST.
KEYWORDS
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 1201)
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization;
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr

```

Db      726  ACACCTGTTGATGTCAGCAAGCATGCTTTGAATGGCAACAATAATTAGACAGACACA 785
QY      743  CCAGAGCCACCACTCTCTGTTTGTTCAGATATTTGAGGATACCTTAAGATCTAGAGTCAAAAT 802
Db      786  CCAGAGCCACCACTCTCTGTTTGTTCAGATATTTGAGGATACCTTAAGATCTAGAGTCAAAAT 845
QY      803  GTTTAAATTCAGAGGGCGTTTCAGATACCTTTTGATCCATATCTTTGATATATACATTTGAGA 862
Db      846  GTTTAAATTCAGAGGGCGTTTCAGATACCTTTTGATCCATATCTTTGATATATACATTTGAGA 905
QY      863  TAAAGCTGCTCAGAGTGTCAACAAGGCATTTGAGCAGTGTGTAAGCCGGAACAGCTTG 922
Db      906  TAAAGCTGCTCAGAGTGTCAACAAGGCATTTGAGCAGTGTGTAAGCCGGAACAGCTTG 965
QY      923  ATGGAAGAAATCGTACAAAGTCAGCAAGTGTAAAGATGTTTCAGCTTCAAGAGGT 982
Db      966  ATGGAAGAAATCGTACAAAGTCAGCAAGTGTAAAGATGTTTCAGCTTCAAGAGGT 1025

RESULT 2
LOCUS   BX354198
DEFINITION BX354198 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
          CDNA clone CSODC016YG21 5-PRIME, mRNA sequence.
ACCESSION BX354198
VERSION   BX354198.1 GI:30367814
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Web : www.genoscope.cns.fr
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4596.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODC016AD11Q1&cluster=4596.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CSODC016AD11Q1.
Location/Qualifiers
1..1006
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODC016YG21"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source
Query Match 34.4%; Score 830; DB 13; Length 1006;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 880; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      30  CGAGGGGATGAGCGAGCGCGCGGTCAGAGTTGAACATACCATAGTTGACAA 89
Db      67  CGAGGGGATGAGCGAGCGCGCGGTCAGAGTTGAACATACCATAGTTGACAA 126
QY      90  AGCTTCTGAATCTTCAGACCCATCAGCTATCAGAAATCAGCTCGGAGTCCGAGGAGT 149
Db      127  AGCTTCTGAATCTTCAGACCCATCAGCTATCAGAAATCAGCTCGGAGTCCGAGGAGT 186

```

```

QY      150  CTCACTGGAGACATGATGATGAGTTCTGCAGCTGGGGTCTGTCTTCAATTGAATGA 209
Db      187  CTCACTGGAGACATGATGATGAGTTCTGCAGCTGGGGTCTGTCTTCAATTGAATGA 246
QY      210  TGTGTCAATACACACATTTCTTTAGGACCAAGTACCTGTGCTGTAGTTTATTGAGTTC 269
Db      247  TGTGTCAATACACACATTTCTTTAGGACCAAGTACCTGTGCTGTAGTTTATTGAGTTC 306
QY      270  ATCTGTACCTGATAAATCAAAACCATCACCACAAAAGGATCAAGCCCTAGTGTATGGCAT 329
Db      307  ATCTGTACCTGATAAATCAAAACCATCACCACAAAAGGATCAAGCCCTAGTGTATGGCAT 366
QY      330  CGTCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGAAGATTTGTCTTAAAGTGGCAACAAAC 389
Db      367  CGTCTCTCCACAGAAAGTTCTTTTCCCATCTCGAAGAAGATTTGTCTTAAAGTGGCAACAAAC 426
QY      390  TCATAGAGTTGAGCTGGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACCT 449
Db      427  TCATAGAGTTGAGCTGGGCTCCAGAAATTTGGCAATACCTGTTTGGCAATGACGACCT 486
QY      450  GCAGTGTCTTAACTACACACACCTCTTCCCAATTTACATGCTATCATGATGAGTACGCGGTAT 509
Db      487  GCAGTGTCTTAACTACACACACCTCTTCCCAATTTACATGCTATCATGATGAGTACGCGGTAT 546
QY      510  AACATGTCATGAGAGGCTTTTGTATGATGTGTACAATGCAAGCACATATTACCCAGGC 569
Db      547  AACATGTCATGAGAGGCTTTTGTATGATGTGTACAATGCAAGCACATATTACCCAGGC 606
QY      570  ACTCAGTAACTCTGGGAGCTTATTAACCAATGTTTGTATCATGATGAGTACGCGGTAT 629
Db      607  ACTCAGTAACTCTGGGAGCTTATTAACCAATGTTTGTATCATGATGAGTACGCGGTAT 666
QY      630  AGCTAGGACCTCCCTTTTGGAAACCAAGAAGATGCCATGAATTCCTTCAATACACTGT 689
Db      667  AGCTAGGACCTCCCTTTTGGAAACCAAGAAGATGCCATGAATTCCTTCAATACACTGT 726
QY      690  TGATGCTATGCAAGAAGCATGCTTGAATGGCAGCAATAAATTAGACACACACCCAGGC 749
Db      727  TGATGCTATGCAAGAAGCATGCTTGAATGGCAGCAATAAATTAGACACACACCCAGGC 786
QY      750  CACCACTCTGTTGTGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAA 809
Db      787  CACCACTCTGTTGTGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAA 846
QY      810  TTGCAAGGCGCTTTTCAGATATCTTTGATCCATATCTTGTATATACATTTGGAGTAAAGGC 869
Db      847  TTGCAAGGCGCTTTTCAGATATCTTTGATCCATATCTTGTATATACATTTGGAGTAAAGGC 906
QY      870  TGCTCAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGC 910
Db      907  TGCTCAGAGTGTCAACAGGCAATTTGGAGCAGTTTGTGAAGC 947

```

```

RESULT 3
BX418054
LOCUS   BX418054 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
DEFINITION CSODF001YG23 5-PRIME, mRNA sequence.
ACCESSION BX418054
VERSION   BX418054.1 GI:30646401
KEYWORDS EST.
SOURCE   Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 952)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```

Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4596.f For
more information about this cluster, see

http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DF001AD12QP1&cluster=4596.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DF001AD12QP1.

FEATURES
source
1..952
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF001YG23"
/tissue.type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: Brain; Vector: pCMVSPORT_6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."

ORIGIN
Query Match 32.5%; Score 785; DB 13; Length 952;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 885; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

23 CGCGCGGCGGAGTGGAGCGAGCGCGCGGTCAGAGTTCGAACATGACCATAG 82
24 CGCGCGGCGGAGGAGTGGAGCGAGCGCGGTCAGAGTTCGAACATGACCATAG 125
83 TTGACAAAGCTTCTGAATCTTCAGACCATCAGCCATCAGATTCAGATTCAGTCCG 142
126 TTGACAAAGCTTCTGAATCTTCAGACCATCAGCCATCAGATTCAGATTCAGTCCG 185
143 AGGAGTCTCAGCTGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 202
186 AGGAGTCTCAGCTGGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 245
203 TGAATGATGTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGT 262
246 TGAATGATGTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGATTCAGT 305
263 CGAGTTCATCTGTACCTGATTAATCAAAACCATCACCACAAAGGATCAAGCCCTAGGT 322
306 CGAGTTCATCTGTACCTGATTAATCAAAACCATCACCACAAAGGATCAAGCCCTAGGT 365
323 ATGSCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAAGTCTTTAAGTGGC 382
366 ATGSCATCGCTCTCCACAGAAAGTCTTTTCCCATCTGAGAAAGTCTTTAAGTGGC 425
383 AACAACTCATAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATG 442
426 AACAACTCATAGTTGGAGCTGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATG 485
443 CAGCACTGCAGTGTAACTACACACACACCTCTTTGCCAATTAATCATGCTATCAATGAC 502
486 CAGCACTGCAGTGTAACTACACACACACCTCTTTGCCAATTAATCATGCTATCAATGAC 545
503 ACTCCAAACATGTCAGAGAGGCTTTTGTATGATGTACATGCAAGCACAATTA 562
546 ACTCCAAACATGTCAGAGAGGCTTTTGTATGATGTACATGCAAGCACAATTA 605
563 CCGAGGCACTCAGTAATCTCGGGAGCTTATTAAACCAATGTTTGTATCAATGAGATGC 622
606 CCGAGGCACTCAGTAATCTCGGGAGCTTATTAAACCAATGTTTGTATCAATGAGATGC 665
623 GCGGTATAGTAGGACCTTCGTTTGGAAACCAAGAGATGCCCATGAAATTCCTTCAAT 682
666 GCGGTATAGTAGGACCTTCGTTTGGAAACCAAGAGATGCCCATGAAATTCCTTCAAT 725

QY 683 ACACGTGTTGATGCTATGCAAGAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACA 742
Db 726 ACACGTGTTGATGCTATGCAAGAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACA 785
QY 743 CCCAGGCCACCACTCTGTTGTCAGATATTTGGAGGATACCTAAGATCTAGAGTCAAT 802
Db 786 CCCAGGCCACCACTCTGTTGTCAGATATTTGGAGGATACCTAAGATCTAGAGTCAAT 845
QY 803 GTTTAAATGCAAGGCGTTTCAGATATCTTTGATCCATATCTTGATATAAATTGGAGA 862
Db 846 GTTTAAATGCAAGGCGTTTCAGATATCTTTGATCCATATCTTGATATAAATTGGAGA 905
QY 863 TAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAG 909
Db 906 TAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAG 952

RESULT 4
BM920091 1063 bp mRNA linear EST 12-MAR-2002
LOCUS AGENCOURT_6706520 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5749963
DEFINITION 5', mRNA sequence.
ACCESSION BM920091
VERSION BM920091.1 GI:19370470
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1063)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLAM12780 row: c column: 20
High quality sequence stop: 760.

FEATURES
source
1..1063
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5749963"
/lab_host="DH10B"
/clone_lib="NIH_MGC_120"
/note="Organ: pooled pancreas and spleen; Vector:
pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
source anonymous pool of spleen and pancreas from 28 YO
male. Library is oligo-dT primed and directionally cloned
(EcoRV site is destroyed upon cloning). Average insert
size 1.5 kb, insert size range 1-2.5 kb. Library is
normalized and enriched for full-length clones and was
constructed by C. Gruber (Invitrogen). Research Genetics
tracking code 025. Note: this is a NIH_MGC Library."

ORIGIN
Query Match 31.3%; Score 756; DB 12; Length 1063;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 806; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 513 ATGTCATCAGAAGCTTTTGTATGATGTGATGATGATGATGATGATGATGATGATGAT 572
Db 17 ATGTCATCAGAAGCTTTTGTATGATGTGATGATGATGATGATGATGATGATGATGAT 76
QY 573 CAGTAATCTGGGAGCTTATTAAACCAATGTTTGTATGATGATGATGATGATGATGATGATGAT 632

Db 77 CAGTAATCTCGGAGCGTTAATAAACCAATGTTTTCATCAATGAGATGGCGGTATAGC 136
QY 633 TAGGCACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGA 692
Db 137 TAGGCACCTCCGTTTGGAAACCAAGAGATGCCCATGAATTCCTTCAATACACTGTTGA 196
QY 693 TGCTATGCAAGAACGATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCAC 752
Db 197 TGCTATGCAAGAACGATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCAC 256
QY 753 CACTCTTGTGTTGTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTAAATTG 812
Db 257 CACTCTTGTGTTGTCAGATATTTGGAGATACCTAAGATCTAGAGTCAAAATGTTAAATTG 316
QY 813 CAAGGCGGTTTCAGATACCTTTCAGTCCATCTTGATATATTAACATTGGAGATAAGGCTGC 872
Db 317 CAAGGCGGTTTCAGATACCTTTCAGTCCATCTTGATATATTAACATTGGAGATAAGGCTGC 376
QY 873 TCAGAGTGTCAACAAAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTTGTATGGAGAAA 932
Db 377 TCAGAGTGTCAACAAAGGCAATGGAGCAGTTTGTGAAGCCGGAACAGCTTGTATGGAGAAA 436
QY 933 CTCGTACAGTGCAGCAAGTGTAAAGATGTTCCAGCTTCAAAGAGGTTCACTATCCA 992
Db 437 CTCGTACAGTGCAGCAAGTGTAAAGATGTTCCAGCTTCAAAGAGGTTCACTATCCA 496
QY 993 TAGATCCTCTAATGTTCTTACACTTCTCTGAACGTTTGTGAAGCTTTCACGGTGGAAA 1052
Db 497 TAGATCCTCTAATGTTCTTACACTTCTCTGAACGTTTGTGAAGCTTTCACGGTGGAAA 556
QY 1053 AATTGCTAAGGATGTGAATACCTCGAGTATCTTGATATTCGGCCATATATGTCACACC 1112
Db 557 AATTGCTAAGGATGTGAATACCTCGAGTATCTTGATATTCGGCCATATATGTCACACC 616
QY 1113 CAACGAGAGCAATGTCTAGCTCTGATGACAGTGTGTCACACTGTTTAAATTG 1172
Db 617 CAACGAGAGCAATGTCTAGCTCTGATGACAGTGTGTCACACTGTTTAAATTG 676
QY 1173 CCATGCTGGCAATTAATCTCTACATAAAGCTAGCAATGGCCCTCTGGTATCAATGAA 1232
Db 677 CCATGCTGGCAATTAATCTCTACATAAAGCTAGCAATGGCCCTCTGGTATCAATGAA 736
QY 1233 TGACTCCATGATCTACAGTGAATTAAGTACGTTCTCAGCCCAACAGCCCTATGTCCT 1292
Db 737 TGACTCCATGATCTACAGTGAATTAAGTACGTTCTCAGCCCAACAGCCCTATGTCCT 796
QY 1293 CTTTATATCAGTCCCATGATGAA 1319
Db 797 CTTTATATCAGTCCCATGATGAA 823

RESULT 5
LOCUS BUI73371
DEFINITION BUI73371 894 bp mRNA linear EST 04-SEP-2002
5', mRNA sequence.
ACCESSION BUI73371
VERSION BUI73371.1 GI:22687355
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 894)
AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM13548 row: g column: 07
High quality sequence stop: 670.
FEATURES
source Location/Qualifiers
1..894
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6174750"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 71"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: Sall; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."

ORIGIN

Query Match 30.8%; Score 744; DB 13; Length 894;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 794; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 109 CCATCAGCCTATCAGATCAGCTGGCAGCTCGGAGCAGTCTCAGCTGGAGACATGGAT 168
Db 1 CCATCAGCCTATCAGATCAGCTGGCAGCTCGGAGCAGTCTCAGCTGGAGACATGGAT 60
QY 169 GCAGTGTCTGCCAGCTGGGCTGCTGCTTCATTTGAATGATGTGTCAAATCAGACACTT 228
Db 61 GCAGTGTCTGCCAGCTGGGCTGCTGCTTCATTTGAATGATGTGTCAAATCAGACACTT 120
QY 229 TCTTTAGCAACGATACCTGGTGTCTAGTTTATTCGAGTTCACTCTGACCTGATATAATCA 288
Db 121 TCTTTAGCAACGATACCTGGTGTCTAGTTTATTCGAGTTCACTCTGACCTGATATAATCA 180
QY 289 AAACCATCACCAAAAGATCAAGCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTT 348
Db 181 AAACCATCACCAAAAGATCAAGCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTT 240
QY 349 CTTTTCCTCATCTGAGAAGATTGCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGG 408
Db 241 CTTTTCCTCATCTGAGAAGATTGCTTAAGTGGCAACAACTCATAGAGTTGGAGCTGG 300
QY 409 CTCAGAAATTTGGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 468
Db 301 CTCAGAAATTTGGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 360
QY 469 CCACCTCTTGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 528
Db 361 CCACCTCTTGGCAATACCTGTTTCCCAATGAGCACTGCACTGTTTAACTACACA 420
QY 529 TTTTGTATGATGTGTACAAATGCAAGCACATATTACCCAGGCACCTCAGTAATCTCGGAC 588
Db 421 TTTTGTATGATGTGTACAAATGCAAGCACATATTACCCAGGCACCTCAGTAATCTCGGAC 480
QY 589 GTTATTTAAACCAATGTTTGTTCATCAATGAGATGGCGGTATAGCTAGGCACCTCGT 648
Db 481 GTTATTTAAACCAATGTTTGTTCATCAATGAGATGGCGGTATAGCTAGGCACCTCGT 540
QY 649 GGAAACCAAGAGATGCCCATGAAATTCCTTCAATACACTGTTGATGCTATGCAAGACGA 708
Db 541 GGAAACCAAGAGATGCCCATGAAATTCCTTCAATACACTGTTGATGCTATGCAAGACGA 600
QY 709 TGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCTCTGTTGTCAG 768
Db 601 TGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCTCTGTTGTCAG 660
QY 769 ATATTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTCAGAGGCGTTTCAGAT 828
Db 661 ATATTGGAGATACCTAAGATCTAGAGTCAAAATGTTTAAATTCAGAGGCGTTTCAGAT 720
QY 829 ACTTTTGCATATCTTGTATATAACATTGGAGATAAGGCTGCTCAGAGTGTCAACAAG 888

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN	Query Match	Best Local Similarity	Score	DB	Length	Indels	Gaps
	Matches 749; Conservative	99.9%; Pred. No. 0;	29.0%;		918;	1;	0;
QY	864	AAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTCAAGCCGGAACAGCTTGA	923				
Db	21	AAAGGCTGCTCAGAGTGTCAACAGGCAATGGAGCAGTTTGTCAAGCCGGAACAGCTTGA	80				
QY	924	TGAGAGAAACTCGTACAGTGCAGCAGTGTAAAGAGATGGTTCAGCTTCAAGAGGTT	983				
Db	81	TGAGAGAAACTCGTACAGTGCAGCAGTGTAAAGAGATGGTTCAGCTTCAAGAGGTT	140				
QY	984	CACATCCATAGATCCCTAAATGTTCTTACACATTTCTCTGAAACGTTTGTGCAATTTTAC	1043				
Db	141	CACATCCATAGATCCCTAAATGTTCTTACACATTTCTCTGAAACGTTTGTGCAATTTTAC	200				
QY	1044	CGGTGAAAAATGCTAAGGATGTGAATACCTCGATATCTTGATATTCGGCCATATAT	1103				
Db	201	CGGTGAAAAATGCTAAGGATGTGAATACCTCGATATCTTGATATTCGGCCATATAT	260				
QY	1104	GTCTCAACCAACGAGAGCCAAATGTCTACGCTCTGTATGTCAGTCTGGTTCACACTGG	1163				
Db	261	GTCTCAACCAACGAGAGCCAAATGTCTACGCTCTGTATGTCAGTCTGGTTCACACTGG	320				
QY	1164	TTTTAATTGCCATGCTGGCCATTAATCTTGCTACATAAAAGCTAGCAATGGCTCTGGTA	1223				
Db	321	TTTTAATTGCCATGCTGGCCATTAATCTTGCTACATAAAAGCTAGCAATGGCTCTGGTA	380				
QY	1224	TCAATGAATGACTCCATTTGATACAGTGTATACAGTGTATAGATCGGTACTCAGCAACAAGC	1283				
Db	381	TCAATGAATGACTCCATTTGATACAGTGTATACAGTGTATAGATCGGTACTCAGCAACAAGC	440				
QY	1284	CTATGTGCTCTTTTATATCAGTGTCCATGATGTGAAATGAGTGTAACTTACTCATCC	1343				
Db	441	CTATGTGCTCTTTTATATCAGTGTCCATGATGTGAAATGAGTGTAACTTACTCATCC	500				
QY	1344	CACCATAGCCGCGCCAGTCTCTCCCGCCCGTCACTCAGTCAAGCGGTTGTCAACAA	1403				
Db	501	CACCATAGCCGCGCCAGTCTCTCCCGCCCGTCACTCAGTCAAGCGGTTGTCAACAA	560				
QY	1404	CAACAGGCTGGCCAGGCTTTATCGGACACAGCTTCCCTCTCACATGATAAAGAAATCC	1463				
Db	561	CAACAGGCTGGCCAGGCTTTATCGGACACAGCTTCCCTCTCACATGATAAAGAAATCC	620				
QY	1464	ACCTCACTTAAATGGAGCTGGACCATTTGAAGACACGCAAGCAGTTCATGTGAGTCC	1523				
Db	621	ACCTCACTTAAATGGAGCTGGACCATTTGAAGACACGCAAGCAGTTCATGTGAGTCC	680				
QY	1524	TACCGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGCTTCAGTCTCTGTCCTCAAA	1583				
Db	681	TACCGGAATTCAGTGTCAACAGGCTAGTCCCTGTTAATGCTTCAGTCTCTGTCCTCAAA	740				
QY	1584	CTGGTCAGTTAATAGGCTCTCAGTGATCCC	1613				
Db	741	CTGGTCAGTTAATAGGCTCTCAGTGATCCC	770				

AUI24746 851 bp mRNA linear EST 01-AUG-2002
 AUI24746 NT2RM4 Homo sapiens cDNA clone NT2RM4000488 5', mRNA
 sequence.
 AUI24746
 AUI24746.1 GI:10949462
 EST.
 Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 851)
 AUTHORS Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S., Saito, K., Yamamoto, J.,
 Nakamura, Y., Nishikawa, T., Nagai, T., Suzuki, Y., Sugano, S. and
 Isogai, T.
 TITLE HRI human cDNA project (Ota, T., Wakamatsu, A., Ozawa, M., Ishii, S.,
 Saito, K., Yamamoto, J., Nakamura, Y., Nishikawa, T., Nagai, T.,
 Suzuki, Y., Sugano, S., Isogai, T.)
 JOURNAL Unpublished (2000)
 COMMENT Contact: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5' - & 3' - end one pass sequencing: Helix
 Research Institute; cDNA library construction: Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.
 FEATURES
 Location/Qualifiers
 1..851
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4000488"
 /cell_type="teratocarcinoma"
 /cell_line="NT2"
 /clone_lib="NT2RM4"
 /note="Vector: pME18SPL3; mRNA from uninduced NT2 neuronal
 precursor cells"

ORIGIN
 Query Match 28.7%; Score 692; DB 9; Length 851;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 692; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1339 CATCCCAACCAAGGTCGCCAGGCTTTATCGGACCAAGCAGTTCCTCTCACATGATAAG 1458
 Db 2 CATCCCAACCAAGGTCGCCAGGCTTTATCGGACCAAGCAGTTCCTCTCACATGATAAG 121
 QY 1399 ACCAACAACCAAGGTCGCCAGGCTTTATCGGACCAAGCAGTTCCTCTCACATGATAAG 1458
 Db 62 ACCAACAACCAAGGTCGCCAGGCTTTATCGGACCAAGCAGTTCCTCTCACATGATAAG 121
 QY 1459 AATCCACTCTAATTAATGGGACTGGACCAATGAAAGACAGCAGGTCCTCATGTCG 1518
 Db 122 AATCCACTCTAATTAATGGGACTGGACCAATGAAAGACAGCAGGTCCTCATGTCG 181
 QY 1519 AGTCCTAAGCGGAAATTCAGTGTCAAGGCTAGTCTGTTAATGCTTCAGTCTCTGTC 1578
 Db 182 AGTCCTAAGCGGAAATTCAGTGTCAAGGCTAGTCTGTTAATGCTTCAGTCTCTGTC 241
 QY 1579 CAAATCTGTCAGTAAATAGTCTCTCAGTGTATCCAGAAATCTCCTAAGAAACAAATAAT 1638
 Db 242 CAAATCTGTCAGTAAATAGTCTCTCAGTGTATCCAGAAATCTCCTAAGAAACAAATAAT 301
 QY 1639 ACATCAGTATTCACACAAAGTTCCTGTCGCGAGTGTGAGTCTCAACCTTAACCTTCAT 1698
 Db 302 ACATCAGTATTCACACAAAGTTCCTGTCGCGAGTGTGAGTCTCAACCTTAACCTTCAT 361
 QY 1699 AGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTCATTCGCA 1758
 Db 362 AGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCATTCATTCGCA 421
 QY 1759 GTACAGTCTACCTCGAACGCAATCTAGCATGTCTAGTCTTCTAGTAAAGTAAACCAAGC 1818
 Db 422 GTACAGTCTACCTCGAACGCAATCTAGCATGTCTAGTCTTCTAGTAAAGTAAACCAAGC 481
 QY 1819 CCGGCGAGTGAATCTCTGCTCCCGCGCGTGTGAATGGGAAATCCAAGCTGAACCTCCAGC 1878

QY	1226	AAATGAATGACTCCATTGTATCTTACCAGTGAATATTAGATCGGTACTCAGCCAAACAGCCT	1281
DB	61	AAATGAATGACTCCATTGTATCTTACCAGTGAATATTAGATCGGTACTCAGCCAAACAGCCT	120
QY	1286	ATGTGCTCTTTTATATCAGTGTCCCATGATCTGAAAAATGGAGTGAACTTATCTCATCCCA	134
CB	121	ATGTGCTCTTTTATATCAGTGTCCCATGATGAAAAATGGAGTGAACTTATCTCATCCCA	180

```
Location/Qualifiers
1. .822
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```

```

/clone="NT2RM4000431"
/cell_type="teratocarcinoma"
/cell_line="NT2"
/clone_lib="NT2RM4"
/preset:Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
precursor cells"

```

ORIGIN	precursor cells"	Query Match	27.5%;	Score 663;	DB 9;	Length 822;
		Best Local Similarity	99.9%;	Pred. No. 0;		
		Matches	713;	Conservative	0;	Mismatches
					1;	Indels
					0;	Gaps
					0;	
QY	856	TTGAGATAAAGCGTCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCCGGAA	915			
Db	1	TTGAGATAAAGCGTCTCAGAGTGTCAACAAGGCATTGGAGCAGTTTGTGAAGCCGGAA	60			
QY	916	CAGCTTCATGGAGAAACTCGTCAAGTCGAGCAAGTGTAAAAGATGGTTCCAGCTTCA	975			
Db	61	CAGCTTCATGGAGAAACTCGTCAAGTCGAGCAAGTGTAAAAGATGGTTCCAGCTTCA	120			
QY	976	AAGAGGTTCACTATCCATAGATCCTCTAATGTTCTTACATCTTCTCTGAAAACGTTTGA	1035			
Db	121	AAGAGGTTCACTATCCATAGATCCTCTAATGTTCTTACATCTTCTCTGAAAACGTTTGA	180			
QY	1036	AAATTTTACC CGTGGAAAAATTTGCTAAGGATGTGAATACCTTCAGTATCTTGATATCGG	1095			
Db	181	AAATTTTACC CGTGGAAAAATTTGCTAAGGATGTGAATACCTTCAGTATCTTGATATCGG	240			
QY	1096	CCATATATGTTCTCAACCCCAACGGAGAGCCAAATGTCTACGCTCTGTATGTCAGTGTCTGTC	1155			
Db	241	CCATATATGTTCTCAACCCCAACGGAGAGCCAAATGTCTACGCTCTGTATGTCAGTGTCTGTC	300			
QY	1156	CACACTGGTTTTAATTTGCCATGTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGC	1215			
Db	301	CACACTGGTTTTAATTTGCCATGTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGC	360			
QY	1216	CTCTGGTATCAAAATGAATGACTCCCATTTGTAATCAACAGTATATTAGATCGGTACTCAGC	1275			
Db	361	CTCTGGTATCAAAATGAATGACTCCCATTTGTAATCAACAGTATATTAGATCGGTACTCAGC	420			
QY	1276	CAACAAGCCTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAATGGAGGTGAATTT	1335			
Db	421	CAACAAGCCTATGTGCTCTTTTATATACAGTCCCATGATGTGAAAATGGAGGTGAATTT	480			
QY	1336	ACTCATCCACCCATAGCCCCCGCCAGTCCCTCTCCCGCCCGCTCATCAGTCAGCGGGTT	1395			
Db	481	ACTCATCCACCCATAGCCCCCGCCAGTCCCTCTCCCGCCCGCTCATCAGTCAGCGGGTT	540			
QY	1396	GTCACACAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTCTCACATGATA	1455			
Db	541	GTCACACAAACAGGCTGCGCCAGGCTTTATCGGACCAACAGCTTCCCTCTCACATGATA	600			
QY	1456	AAGATCCACCTCATTAAATGGGACTGGACCAATTTGAAAGACAGCCAGCAGTTCCATG	1515			
Db	601	AAGATCCACCTCATTAAATGGGACTGGACCAATTTGAAAGACAGCCAGCAGTTCCATG	660			
QY	1516	TCGAGTCTTAAGGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTTAATGCTTCA	1569			
Db	661	TCGAGTCTTAAGGGAAATTCAGTGTCAACAGGGCTAGTCTCTGTTAATGCTTCA	714			

RESULT 11
BM563326
LOCUS
DEFINITION
1088 bp mRNA linear EST 20-FEB-2002
AGENCOURT_6589448 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5441601
5', mRNA sequence.
ACCESSION
BM563326
VERSION
BM563326.1 GI:18810154
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

```

FEATURES
  1 (bases 1 to 1098)
  NIH-MGC http://mgc.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: ATCC
  cDNA Library Preparation: Rubin Laboratory
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: LLM1916 row: c column: 10
  High quality sequence stop: 611.
  Location/Qualifiers
    1..1098
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:5441601"
    /tissue_type="astrocytoma grade IV, cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH_MGC_98"
    /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
    EcoRI; cDNA made by oligo-dT priming. Directionally-
    cloned into EcoRI/XhoI sites using the following 5',
    adaptor: GGACAGAG(G). Library constructed by Ling Hong
    in the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies).
    Note: this is a NIH MGC Library."

```

ORIGIN

Query Match	25.0%; Score 604; DB 12; Length 1088;
Best Local Similarity	99.8%; Pred. No. 5.9e-298;
Matches 654; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	133 GCAGCTCCGAGGCGAGTCTCACTCGAGACATGGATGCAGGTTCTGCGACGCTGGGFGCT 192
Db	1 GCGAGCTCCGAGGCGAGTCTCACTCGAGACATGGATGCAGGTTCTGCGACGCTGGGFGCT 60
QY	193 GTGTCTTCAATGAATGAATGTCAAAATCACACACTTTCTTTAGGACAGGTAACCTGGTGCT 252
Db	61 GTGTCTTCAATGAATGAATGTCAAAATCACACACTTTCTTTAGGACAGGTAACCTGGTGCT 120
QY	253 GTAGTTTATTCGAGTTTATCTGTACTGTATAATCAAAACCATCACCAAAAAGGATCAA 312
Db	121 GTAGTTTATTCGAGTTTATCTGTACTGTATAATCAAAACCATCACCAAAAAGGATCAA 180
QY	313 GCGCTAGGTGATGGCATCGCTCCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTGT 372
Db	181 GCGCTAGGTGATGGCATCGCTCCTCCACAGAAAGTCTTTTCCCATCTGAGAAGATTGT 240
QY	373 CTTAAGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 432
Db	241 CTTAAGTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGT 300
QY	433 TTTGGCAATGAGGACATCGAGTGTTTAACTTACCTACACACACCTCTTGGCCAAATTACATGCTA 492
Db	301 TTTGGCAATGAGGACATCGAGTGTTTAACTTACCTACACACACCTCTTGGCCAAATTACATGCTA 360
QY	493 TCACATGAACACTCCAAAACATGTCATGCAGAGGCTTTTGTATCATGTGTACAATGCAA 552
Db	361 TCACATGAACACTCCAAAACATGTCATGCAGAGGCTTTTGTATCATGTGTACAATGCAA 420
QY	553 GCACATATTACCGAGGCACTCAGTAATCCTGGGACGTTATTAACACCAATGTTTGTGATC 612
Db	421 GCACATATTACCGAGGCACTCAGTAATCCTGGGACGTTATTAACACCAATGTTTGTGATC 480
QY	613 AATCAGATGCGGCGTATAGCTTAGGCACTCCGTTTGTGAAAACCAAGAAGATGCCCATGAA 672
Db	481 AATCAGATGCGGCGTATAGCTTAGGCACTCCGTTTGTGAAAACCAAGAAGATGCCCATGAA 540

673	TTCTTCAATACACTGTTGATCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTA	732
541	TTCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTA	600
733	GCAGACACACCCAGGCCACCACTCTTGTGTTGTCAGATATTGGAGGATACCTTAA	787
601	GCAGACACACCCAGGCCACCACTCTTGTGTCAGATATTGGAGGATACCTTAA	655

RESULT 12

BQ778581
LOCUS
DEFINITION
BQ778581 595 bp mRNA linear EST 26-JUL-2002
il29f12.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6031630 5' ,
mRNA sequence.

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
BQ778581	BQ778581.1	GI:21987053 EST.	Homo sapiens (human)	Homo sapiens
				Homo sapiens
				Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1. (bases 1 to 595)

Melton, D., Brown, J., Keny, G., Permut, A., Lee, C., Kaestner, K.,
Lewin, C., Brown, J., M., Brestelli, J., Gradwohl, G., Clifton, S.,
Leishka, I., Seacore, M., Pape, D., Wylie, T., Martin, J., Blustain, A.,
Hall, L., Maria, M., Ritter, E., Ronko, I., Bennett, J.,
Schmitt, A., Theising, B., R., McCann, R., Cole, R., Tsagarisvili, R.,
Gardnas, M., Gibbons, M., Williams, T. Jackson, Y. and Bowers, Y.

TITLE	Endocrine Pancreas Consortium
JOURNAL	Unpublished (2000)
COMMENT	Other ESTs: il29f12.x1 Contact: Douglas Maltou, Klaus H. Kaestner, & Hiroshi Inoue Endocrine Pancreas Consortium Endocrine Pancreas Consortium Harvard University, Howard Hughes Medical Institute Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinooue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 485.

```

Location/Qualifiers
1. 595
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="caxon:9606"
/clone="IMAGE:6031630"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1
NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
314-362-1976. Fax: 314-747-2692."

```

```

ORIGIN
      24.3%; Score 595; DB 13; Length 595;
Query Match
Best Local Similarity 100.0%; Pred. No. 2.9e-289;
Matches 595; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1371 CCGCCCCGTCATCAGTCAGCGGGTTTCTACCAACAAACAGGCTGCCCGAGGCTTTATCGG 143
      11 CCGCCCCGTCATCAGTCAGCGGGTTTCTACCAACAAACAGGCTGCCCGAGGCTTTATCGG 70

```

[illegible][illegible]

Contact: Robb-remail.nih.gov
 Email: cgbps-remail.nih.gov
 Oligo-dt track not found, Not I site shown in beginning of sequence
 is likely internal to the message. Issue Procurement: Louis M.
 Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab
 cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone Distribution: MSC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 seq primer: M13 Forward
 POLYA-No.

FEATURES	source	Location/Qualifiers
		1..557
		/organism="Homo sapiens"
		/mol_type="mRNA"
		/db_xref="taxon:9606"
		/clone="IMAGE:3066811"

RESULT 15
 AW298129/c
 LOCUS
 DEFINITION
 UI-R-BWO-ajs-e-07-0-UI.sl NCI CGAP_Sub6 Homo sapiens cDNA clone
 IMAGE:2733012 3', mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 538)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Oligo-dt track not found. Not 1 site shown in beginning of sequence
 is likely internal to the message. cDNA library preparation: M.B.
 Soares Lab Clone distribution: NCI-CGAP clone distribution
 information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=NO.

FEATURES
 Location/Qualifiers
 1..538
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2733012"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pTV3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not 1; Site 2: Eco RI; NCI CGAP Sub6
 is a substracted library derived from BW, which consists of
 a mixture of four normalized libraries: NCI CGAP Brn50,
 NCI CGAP Lu13, NCI CGAP Ov18, GBC1. The NCI CGAP Sub6
 library had 7 million recombinants. A single-stranded DNA
 preparation of BW was used as a tracer in a subtractive
 hybridization with a driver comprising: the IMAGE pool
 (NCI CGAP_Kid3 pool 1 LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE ClonesIDs 1322376-1323911,
 1456008-1456775, 1500552-1502855); NCI CGAP Kids pool 1
 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE ClonesIDs
 1323912-1325831, 1471368-1472903, 1492104-1493255);
 NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3851-3854 (IMAGE
 ClonesIDs 1414920-1417991, 1520904-1522439); NCI CGAP_GC4
 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
 ClonesIDs 1257096-1258631, 1469064-1470983,
 1475592-1476743); NCI CGAP Pr22 pool 1 LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE ClonesIDs 985608-986759,
 1101192-1101959, 1217928-1220615); NCI CGAP_Co10 pool 1
 LLAM 2644-2653, 2871-2872 (IMAGE ClonesIDs
 1057416-1061255, 1144584-1145351). (50% of the driver
 population), plus a pool of 3,840 arrayed clones from
 NCI CGAP Sub1 (IMAGE ClonesIDs 2708616-2710535) and
 NCI CGAP Sub2 (IMAGE ClonesIDs 2710536-2712455) (20% of
 the driver population), plus a pool of 11,136 clones from
 NCI CGAP Sub3 (IMAGE ClonesIDs 2712456-2723591) (30% of
 the driver population). Subtraction was performed as
 previously described (Honald, Lennon & Soares (1996):
 Normalization and Subtraction: Two Approaches To
 Facilitate Gene Discovery. Genome Research 6, 791-806.
 TAG TISSUE=brain
 TAG LIB=NCI CGAP_Brn50
 TAG_SEQ=TTTGG"

ORIGIN

Query Match 22.3%; Score 538; DB 10; Length 538;
 Best Local Similarity 100.0%; Pred. No. 3.8e-264;

Matches 538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1834	TCCTCCAGCCCGTGAATGAATGGCAATCAAGCTGAATCCAGCGTGTGGTGCCTAT	1893
Db	538	TCCTCCAGCCCGTGAATGAATGGCAATCAAGCTGAATCCAGCGTGTGGTGCCTAT	479
QY	1894	GGCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGAGATGGG	1953
Db	478	GGCGCGAGTCTCTGAGGACTCTGACGAGGAGTCAAGGGGCTGGGCAAGAGATGGG	419
QY	1954	ATTGGTACGATTGTGAGTCCCACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCAC	2013
Db	418	ATTGGTACGATTGTGAGTCCCACTCTCCCGCCAAAGATGCCGAAGATGAGGAGGCAC	359
QY	2014	CGCAGCAGCTTCAGAAACCCATGACCTTAACGGTGTCTAATAGTCACAGCAGCAG	2073
Db	358	CGCAGCAGCTTCAGAAACCCATGACCTTAACGGTGTCTAATAGTCACAGCAGCAG	299
QY	2074	GACCCGAAAGAAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCTGCCCTG	2133
Db	298	GACCCGAAAGAAAACGGCTAGCGCTGATGGTCCAGCTGCCAAGGCCAGCTGCCCTG	239
QY	2134	CACTCAGAAAATCCCTTTGCTTAAGGCAACGGTCTTCCTGGAAGTTGATGCTGCTCT	2193
Db	238	CACTCAGAAAATCCCTTTGCTTAAGGCAACGGTCTTCCTGGAAGTTGATGCTGCTCT	179
QY	2194	TTGCTGTCTCTCCCAAGACACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAA	2253
Db	178	TTGCTGTCTCTCCCAAGACACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAA	119
QY	2254	GGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGGCCCTCCCGAGGACCGGAC	2313
Db	118	GGCTCGACGATGAATGAGTGCACCTGGAGCAGAGAGGGGCCCTCCCGAGGACCGGAC	59
QY	2314	GCGAGCCTCAGCTGCGAGCCCGCCCGCGAATCCTCTGGAGAGCAGATGCGGCGG	2371
Db	58	GCGAGCCTCAGCTGCGAGCCCGCCCGCGAATCCTCTGGAGAGCAGATGCGGCGG	1

RESULT 16
 AW500503
 LOCUS
 DEFINITION
 UI-HF-BN0-akj-a-03-0-UI.r1 NIH_MGC_50 Homo sapiens cDNA clone
 IMAGE:3077260 5', mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 542)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab
 cDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 Location/Qualifiers
 1..542
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3077260"
 /tissue_type="lymph"

FEATURES
 source


```

/cell_type="germinal center B cells"
/cell_line="WGC85"
/lab_host="DH10B (LT1)"
/clone_lib="NIH MGC_50"
/notes="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(3.5-4.4kb). Directionally cloned. Cells provided by
Louis M. Staudt, Ph.D. Library preparation by Maria de
Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

ORIGIN
Query Match      22.2%; Score 535; DB 10; Length 542;
Best Local Similarity 100.0%; Pred. No. 1.3e-262;
Matches 535; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1265 CGGTACTACGCAACAGCTATGCTCTTTATATACAGTCCCATGATGAAAAATG 1324
Db      8 CGGTACTACGCAACAGCTATGCTCTTTATATACAGTCCCATGATGAAAAATG 67

QY 1325 GAGGTGAATCTACTATCCACCATAGCCCGGCGAGTCCCTCCCGCCCGGTATCA 1384
Db      68 GAGGTGAATCTACTATCCACCATAGCCCGGCGAGTCCCTCCCGCCCGGTATCA 127

QY 1395 GTACGCGGTTGTACCAACAAACAGGTGCGCCAGGCTTTATCGACACAGCTTCCT 1444
Db      128 GTACGCGGTTGTACCAACAAACAGGTGCGCCAGGCTTTATCGACACAGCTTCCT 187

QY 1445 CTCACATGATAAAGATCCACCTCACTTAATGAGGACTGACCATGAAAGACAGCCAA 1504
Db      188 CTCACATGATAAAGATCCACCTCACTTAATGAGGACTGACCATGAAAGACAGCCAA 247

QY 1505 GCAGTTCCATGCGAGTCTTAACGGGAATTCAGAGTCAACAGGCTAGTCTGTTAATG 1564
Db      248 GCAGTTCCATGCGAGTCTTAACGGGAATTCAGAGTCAACAGGCTAGTCTGTTAATG 307

QY 1565 CTTGAGCTTCTGTCCAAACTGGTGCAGTAAATAGTCTCTCAGTGATCCAGAAATCCTTA 1624
Db      308 CTTGAGCTTCTGTCCAAACTGGTGCAGTAAATAGTCTCTCAGTGATCCAGAAATCCTTA 367

QY 1625 AGAACAAATAATACATCAGTATTCACAGAGTTCAGTTCGCTGTCGCGAGTGAGTCTC 1684
Db      368 AGAACAAATAATACATCAGTATTCACAGAGTTCAGTTCGCTGTCGCGAGTGAGTCTC 427

QY 1685 AACCTAACCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCA 1744
Db      428 AACCTAACCTTCATAGTAAATCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCTACCA 487

QY 1745 TTACCAATTCGAGTACAGTCTACCTCGAAGCATCTACGATGTCAGTTTCTAG 1799
Db      488 TTACCAATTCGAGTACAGTCTACCTCGAAGCATCTACGATGTCAGTTTCTAG 542

```

```

RESULT 17
AW976266/c
LOCUS      AW976266
DEFINITION EST388375 MAGE resequences, MAGN Homo sapiens cDNA, mRNA sequence.
ACCESSION AW976266
VERSION    AW976266.1 GI:8167492
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 686)
Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspar,D., Gay,C.,
Holt,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,R.J. and
Quackenbush,J.
Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

```

```

Tel: 301 838 3528
Fax: 301 838 0208
Email: johnq@tigr.org
Plate: 358
Seq primer: Forward.
Location/Qualifiers
source      1..686
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGN"
/notes="Vector: pBluescriptSkm"

FEATURES
ORIGIN
Query Match      22.1%; Score 534; DB 10; Length 686;
Best Local Similarity 99.8%; Pred. No. 4.5e-262;
Matches 584; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1755 TGCAGTACAGTCTACCTCGAAGCGATCTACGATGTCAGTTTCTAGTAAAGTAAACAAACC 1814
Db      585 TGCAGTACAGTCTACCTCGAAGCGATCTACGATGTCAGTTTCTAGTAAAGTAAACAAACC 526

QY 1815 GATCCCCCGCAGTGAATCTCTGCTCCAGCCCTGATGATGAATGCAATCCAAAGTGAATC 1874
Db      525 GATCCCCNGCAGTGAATCTCTGCTCCAGCCCTGATGATGAATGCAATCCAAAGTGAATC 466

QY 1875 CAGCGTGTGTGTCCTATGCGCGCAGTCTCTCTGAGGACTCTGACGAGGAGTCAAAAGG 1934
Db      465 CAGCGTGTGTGTCCTATGCGCGCAGTCTCTCTGAGGACTCTGACGAGGAGTCAAAAGG 406

QY 1935 GCTGGGCAAGGAGATGGAGTTGGTACGATGTGAGTCTCCACTCTCCCGGCAAGATGC 1994
Db      405 GCTGGGCAAGGAGATGGAGTTGGTACGATGTGAGTCTCCACTCTCCCGGCAAGATGC 346

QY 1995 CGAAGTACAGGAGGCACTCCGCGAGCTTCAAGAACCCATGACCCCTAAACCGTGTAA 2054
Db      345 CGAAGTACAGGAGGCACTCCGCGAGCTTCAAGAACCCATGACCCCTAAACCGTGTAA 286

QY 2055 TAGTCACAGCAGCAGTACCTCCGAGAAAGAAAGAAAGGCTGATGTGCCAGCTG 2114
Db      285 TAGTCACAGCAGCAGTACCTCCGAGAAAGAAAGAAAGGCTGATGTGCCAGCTG 226

QY 2115 CCAAGGCCAGCTCCCTGCGACTCAGAAATCCCTTTGCTAAGGAAACCGTCTTCCTGG 2174
Db      225 CCAAGGCCAGCTCCCTGCGACTCAGAAATCCCTTTGCTAAGGAAACCGTCTTCCTGG 166

QY 2175 AAAGTTGATGCTGCTCTCTTTGCTGCTCTCCAGAACACAAATCTTTAGAGACCTTCAG 2234
Db      165 AAAGTTGATGCTGCTCTCTTTGCTGCTCTCCAGAACACAAATCTTTAGAGACCTTCAG 106

QY 2235 GCTTAGCAACAACTGAAAGCTCGACGATGAATGATGCACTCGGAGCAGAGGGG 2294
Db      105 GCTTAGCAACAACTGAAAGCTCGACGATGAATGATGCACTCGGAGCAGAGGGG 46

QY 2295 CCCTCCCGAGGACCGCGAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAG 2339
Db      45 CCCTCCCGAGGACCGCGAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAG 1

```

```

RESULT 18
AW976218/c
LOCUS      AW976218
DEFINITION w63d05.x1 NCI_CGAP Pr22 Homo sapiens cDNA clone IMAGE:246009 3',
mRNA sequence.
ACCESSION AW976218
VERSION    AW976218.1 GI:5675088
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 533)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

```

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsr@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.biol.llnl.gov/hbrp/image/image.html
Insert length: 863 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 490.

```

FEATURES
    source
        1. .533
            Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="rRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:2460009"
                /sex="male"
                /tissue_type="normal prostate"
                /lab_host="DH10B"
                /lab_lib="NCI CGAP Pr22"
                /note="Organ: Prostate; Vector: pT7T3D-Pac (Pharmacia)
                    with a modified polylinker; 1st strand cDNA was prepared
                    from normal prostate bulk tissue, and was then primed with
                    a Not I - oligo(dT) primer. Double-stranded cDNA was
                    ligated to Eco RI adaptors (Pharmacia), digested with Not
                    I and cloned into the Not I and Eco RI sites of the
                    modified pT73 vector. Library is normalized, and was
                    constructed by Bento Soares and M. Fatima Bonaldo."

```

ORIGIN	Query Match	22.1%	Score 533	DB 9	Length 533	
	Best Local Similarity	100.0%	Pred. No. 1.4e-261			
	Matches 533	Conservative 0	Mismatches 0	Indels 0	Gaps 0	Gaps 0
QY	1840	CAGCCCTGATGAATGCGCAATCCAAGCTGAATCCAGCTGAACTCCAGCGTCTGTGTGCGCCATATGGCGCC	1899			
DB	533	CAGCCCTGATGAATGCGCAATCCAAGCTGAATCCAGCTGAACTCCAGCGTCTGTGTGCGCCATATGGCGCC	474			
QY	1900	GAGTCCTCTCAGAGACTCTGACGAGGAGTCAAAAGGGGTGGGCAAGGAGAAATGGGATTGGT	1959			
DB	473	GAGTCCTCTCAGGACTCTGACGAGGAGTCAAAAGGGGTGGGCAAGGAGAAATGGGATTGGT	414			
QY	1960	ACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGCAC	2019			
DB	413	ACGATTGTGAGCTCCCACTCTCCCGGCCAAGATGCCGAAGATGAGGAGGCCACTCCGCAC	354			
QY	2020	GAGCTTCAGAACCCATGACCCCTAAACGGTGCTAATAGTCAGACACGGACAGTGACCCG	2079			
DB	353	GAGCTTCAGAACCCATGACCCCTAAACGGTGCTAATAGTCAGACACGGACAGTGACCCG	294			
QY	2080	AAAGAAACGGCTTAGCGCCTGATGTGCCAGCTGCCAAGGCCAGCCTGCCCTGCACCTCA	2139			
DB	293	AAAGAAACGGCTTAGCGCCTGATGTGCCAGCTGCCAAGGCCAGCCTGCCCTGCACCTCA	234			
QY	2140	GAAATCCCTTTGTAAAGGCAACCGGTCTTCTCGAAAGTTGATGCTGTCTCTTTGCTG	2199			
DB	233	GAAATCCCTTTGTAAAGGCAACCGGTCTTCTCGAAAGTTGATGCTGTCTCTTTGCTG	174			
QY	2200	TCTCTCCAGAAGACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTCG	2259			
DB	173	TCTCTCCAGAAGACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTCG	114			
QY	2260	ACGGATGAATGAGTGCACCTGGACACAGAGGGGCCCTCCCGAGGACCGGACGCCGAG	2319			
DB	113	ACGGATGAATGAGTGCACCTGGACACAGAGGGGCCCTCCCGAGGACCGGACGCCGAG	54			
QY	2320	CCTCAGCCTGGCAGCCCGCCCGCAATCCCTTGGAGGAGCCAGATGCGGGGCGC	2372			

Db 53 CCTCAGCTGGCAGCCCCCGCCGAATCCCTGGAGGACGATGGCGCGC 1

RESULT 19
 AI738779/c
 LOCUS
 DEFINITION

AI738779 553 bp mRNA linear EST 21-DEC-1999
 wi36e04.x1 NCI_CGAP_Col16 Homo sapiens cDNA clone IMAGE:2392350 3',
 mRNA sequence.

ACCESSION
 VERSION
 KEYWORDS
 SOURCE

AI738779
 AI738779.1 GI:5190760
 EST.
 Homo sapiens
 Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 553)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

REFERENCE
 AUTHORS
 TITLE

JOURNAL
 COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapsb@email.nih.gov
 Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 661 Std Error: 0.00
 Seq primer: -40UP from Gibco
 High quality sequence stop: 507.
 Location/Qualifiers
 1. 553
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2392350"
 /tissue_type="colon tumor, RER+"
 /lab_host="DH10B"
 /clone_lib="NCI_CGAP Col6"
 /note="Organ: colon; Vector: p7T73D-Pac (Pharmacia) with a
 modified polylinker; Site 1: Not 1; Site 2: Eco RI;
 plasmid DNA from the normalized library NCI_CGAP Col10 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (cloneIDs 1057416-1061255, and 1144584-1145351).
 Subtraction by Bento Soares and M. Fatima Bonaldo. "

FEATURES
 source

ORIGIN	Query Match	22.1%	Score 533	DB 9	Length 553	Best Local Similarity	100.0%	Pred. No. 1.4e-261	Matches 533	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	1840	CAGCCCGTGATGTAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGGTGCCTATGGCGCC	1895			CAGCCCGTGATGTAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGGTGCCTATGGCGCC	474						
Db	533	CAGCCCGTGATGTAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGGTGCCTATGGCGCC	474			CAGCCCGTGATGTAATGGCAAAATCCAAAGCTGAACTCCAGCGTGTGGTGCCTATGGCGCC	474						
Qy	1900	GAGTCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGGAGAAATGGGATTTGGT	1959			GAGTCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGGAGAAATGGGATTTGGT	414						
Db	473	GAGTCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGGAGAAATGGGATTTGGT	414			GAGTCCTCTGAGGACTCTGACGAGGAGTCAAAGGGCTGGGCAAGGAGAAATGGGATTTGGT	414						
Qy	1960	ACGATTGTGAGCTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGCAC	2015			ACGATTGTGAGCTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGCAC	354						
Db	413	ACGATTGTGAGCTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGCAC	354			ACGATTGTGAGCTCCCACTCTCCCGGCCAAGATCCGAAGATGAGGAGGCCACTCCGCAC	354						
Qy	2020	GAGCTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCGACAGCGACAGTGAACCCG	2071			GAGCTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCGACAGCGACAGTGAACCCG	294						
Db	353	GAGCTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCGACAGCGACAGTGAACCCG	294			GAGCTTCAAGAACCCATGACCCCTAAACGGTCTAATAGTCGACAGCGACAGTGAACCCG	294						

QY 2080 AAAGAAACGGCTAGCGCTGATGTCAGTGCAGTGCAGAGCCAGCGCTGCCCTGCACTCA 2139
 DB 293 AAAGAAACGGCTAGCGCTGATGTCAGTGCAGTGCAGAGCCAGCGCTGCCCTGCACTCA 234
 QY 2140 GAAATCCCTTTGCTTAAGGCAACGGCTCTTCTGGAAGCTTGATGCTCTCTTTGCTG 2199
 DB 233 GAAATCCCTTTGCTTAAGGCAACGGCTCTTCTGGAAGCTTGATGCTCTCTTTGCTG 174
 QY 2200 TCTCTCCAGAGACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTCG 2259
 DB 173 TCTCTCCAGAGACAAATCTTAGAGACCTTCAGGCTTAGCAACAACTGAAAGGCTCG 114
 QY 2260 ACGGATGAAATCAAGTGCACCTGAGCAGAGAGGGCCCTCCGAGAGCCGCGACGCCGAG 2319
 DB 113 ACGGATGAAATCAAGTGCACCTGAGCAGAGAGGGCCCTCCGAGAGCCGCGACGCCGAG 54
 QY 2320 CTTAGCTGGCAGCCCGCCCGCGAATCCCTGGAGGAGCCAGATGCGGCCGC 2372
 DB 53 CTTAGCTGGCAGCCCGCCCGCGAATCCCTGGAGGAGCCAGATGCGGCCGC 1

 RESULT 20
 CFI35585
 LOCUS
 DEFINITION
 UI-HF-BN0-amg-h-08-0-UI.r1 NIH MGC_50 Homo sapiens cDNA clone
 IMAGE:3090278 5', mRNA sequence.
 CFI35585
 CFI35585.1 GI:33251029
 EST.
 Homo sapiens
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 722)
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 889548
 Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRP, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Louis Staudt
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/humanfl.html
 Seq primer: pYX-5.
 Location/Qualifiers
 1..722
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3090278"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /clone_lib="NIH MGC 50"
 /note="Vector: p7T3-Pac; Site 1: NotI; Site 2: Eco RI;
 Constructed from size fractionated cytoplasmic mRNA
 (3.5-4.4kb). Directionally cloned. Cells provided by
 Louis M. Staudt, Ph.D. Library preparation by Maria de
 Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

Query Match 21.5%; Score 518; DB 14; Length 722;
 Best Local Similarity 100.0%; Pred. No. 7.4e-254;
 Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1118 GAGAGCCCAATTCCTACCGTCTTTGATGTCAGTGCCTGGTCCACACTGGTTTAAATGCAATG 1177
 DB 62 GAGAGCCCAATTCCTACCGTCTTTGATGTCAGTGCCTGGTCCACACTGGTTTAAATGCAATG 121
 QY 1178 CTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTATCAAAATGAATGACT 1237
 DB 122 CTGGCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTATCAAAATGAATGACT 181
 QY 1238 CCAATTGTATCTACCAAGTATATAGATCGGTACTCAGCCAAAGCCTATGTGCTCTTTT 1297
 DB 182 CCAATTGTATCTACCAAGTATATAGATCGGTACTCAGCCAAAGCCTATGTGCTCTTTT 241
 QY 1298 ATATCAGGTCCCATCATGTGAAATATGGAGGTGAACCTTACTCATCCACCCATAGCCCG 1357
 DB 242 ATATCAGGTCCCATCATGTGAAATATGGAGGTGAACCTTACTCATCCACCCATAGCCCG 301
 QY 1358 GCCAGTCTCTCCCGCCCGCGTCATCAGTCAGCGGGTTGTCAACCAAAACAGGCTGGCG 1417
 DB 302 GCCAGTCTCTCCCGCCCGCGTCATCAGTCAGCGGGTTGTCAACCAAAACAGGCTGGCG 361
 QY 1418 CAGGCTTTATCGGACCACAGCTTCCCTCTCATGATATAAGAAATCCACTCTCTTAAATG 1477
 DB 362 CAGGCTTTATCGGACCACAGCTTCCCTCTCATGATATAAGAAATCCACTCTCTTAAATG 421
 QY 1478 GGACTGGACCAATGAAAGACACAGCCCAAGCAGTTCCTCGAGTCCGTAACGGGAATTCGA 1537
 DB 422 GGACTGGACCAATGAAAGACACAGCCCAAGCAGTTCCTCGAGTCCGTAACGGGAATTCGA 481
 QY 1538 GTGTCAACAGGGCTAGTCTCTTAAATGCTTCTGCTTCTGTCCTCAAACTGGTCAGTTAATA 1597
 DB 482 GTGTCAACAGGGCTAGTCTCTTAAATGCTTCTGCTTCTGTCCTCAAACTGGTCAGTTAATA 541
 QY 1598 GGTCTCTAGTATCCAGAACATCTTAAGAAACAAAAA 1635
 DB 542 GGTCTCTAGTATCCAGAACATCTTAAGAAACAAAAA 579

 RESULT 21
 CFI35585
 LOCUS
 DEFINITION
 ir68901.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6607849 3',
 mRNA sequence.
 CFI35585
 CFI35585.1 GI:26675619
 EST.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 514)
 Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J.,
 Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R.,
 Williams, T., Jackson, Y. and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@bioh.harvard.edu
 Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Dr. Hiroshi Inoue

(hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 407.
Location/Qualifiers
1. 514
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6607849"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1: NotI; Site 2: XhoI; cDNA made by oligo-dT priming. Size selected on agarose gel. Average insert size ~1kb. 5' XhoI site was destroyed after directional cloning. Amplified once. Contact information: Hiroshi Inoue, MD, Metabolism Div. (Alan Permutt Lab), Washington University School of Medicine, Box 8127, 660 South Euclid Ave., St. Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel: 314-362-1916, Fax: 314-747-2692."

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 771)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11502 row: d column: 02
High quality sequence start: 2
High quality sequence stop: 766.
Location/Qualifiers
1. 771
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5200465"
/lab_host="DH10B"
/clone_lib="NIH MGC 114"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.5 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 019. Note: this is a NIH_MGC Library."

FEATURES
source

REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
ORIGIN	
Query Match	21.3%; Score 514; DB 14; Length 514;
Best Local Similarity	100.0%; Pred. No. 7.8e-252; Indels 0; Gaps 0;
Matches 514; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1823 GCAGTGAATCTCTCCAGCCGCGTATGAATGGCAAAATCCAAAGCTCAATCCAGCGTGC 1882
Db	514 GCAGTGAATCTCTCCAGCCGCGTATGAATGGCAAAATCCAAAGCTCAATCCAGCGTGC 455
QY	1883 TGGTGCCTATGCGCGGAGTCTCTGAGACTCTGACGAGGAGTCAAGGGGCTGGCA 1942
Db	454 TGGTGCCTATGCGCGGAGTCTCTGAGACTCTGACGAGGAGTCAAGGGGCTGGCA 395
QY	1943 AGGAGATGGATTGGTACGATTGTGAGCTCCCACTCTCCCGCCCAAGATGCCGAAGATG 2002
Db	394 AGGAGATGGATTGGTACGATTGTGAGCTCCCACTCTCCCGCCCAAGATGCCGAAGATG 335
QY	2003 AGGAGGCCACTCCGCGAGCTTTCAAGAACCCATGACCTTAAACGGTGTCTAATAGTCAG 2062
Db	334 AGGAGGCCACTCCGCGAGCTTTCAAGAACCCATGACCTTAAACGGTGTCTAATAGTCAG 275
QY	2063 ACAGGACAGTACCCGAAGAAAACGGCTAGCGCTGATGTGCGAGTGCAGTGCAGGCC 2122
Db	274 ACAGGACAGTACCCGAAGAAAACGGCTAGCGCTGATGTGCGAGTGCAGTGCAGGCC 215
QY	2123 AGCTGCGCTGCACTCAGAAAATCCCTTTCTGAAGCAAAACGGTCTTCTCTGGAAGTTGA 2182
Db	214 AGCTGCGCTGCACTCAGAAAATCCCTTTCTGAAGCAAAACGGTCTTCTCTGGAAGTTGA 155
QY	2183 TGGCTGCTCTCTGCTGTCTCTCCAGAGCAAAATCTTAGAGACCTTCAGGGTTAGCA 2242
Db	154 TGGCTGCTCTCTGCTGTCTCTCCAGAGCAAAATCTTAGAGACCTTCAGGGTTAGCA 95
QY	2243 ACAACTGAAGGCTCAGCGATGAATGAGTGCACCTGGAGCAGAGGGGCGCTCCCG 2302
Db	94 ACAACTGAAGGCTCAGCGATGAATGAGTGCACCTGGAGCAGAGGGGCGCTCCCG 35
QY	2303 AGGACCGGAGCGCGAGCTCAGCGCTGGCAGCCC 2336
Db	34 AGGACCGGAGCGCGAGCTCAGCGCTGGCAGCCC 1
RESULT 22	
LOCUS	BI755971 771 bp mRNA linear EST 25-SEP-2001
DEFINITION	63030302P1 NIH_MGC_114 Homo sapiens cDNA clone IMAGE:5200465 5', mRNA sequence.
ACCESSION	BI755971
VERSION	BI755971.1 GI:15747549
KEYWORDS	EST.
SOURCE	Homo sapiens (human)

```

QY      1312 GATGTGAAAAATGGAGGTGAACCTTACTCATCCCA 1345
      |||||||
Db       622 GATGTGAAAAATGGAGGTGAACCTTACTCATCCCA 655

RESULT 23
AI472304/c
LOCUS   AI472304
DEFINITION tJ87a05.xl Soares NSF F8 9W OT PA_P_S1 Homo sapiens cDNA clone
IMAGE:2148464 3', mRNA sequence.
ACCESSION AI472304
VERSION    AI472304.1 GI:4334394
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 488)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
          Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          This clone is available royalty-free through LLNL ; contact the
          IMAGE Consortium (info@image.llnl.gov) for further information.
          Insert Length: 1221 Std Error: 0.00
          Seq primer: -40UP from Gibco
          High quality sequence stop: 464.
          Location/Qualifiers
            1..488
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone="IMAGE:2148464"
              /lab_host="DH10B"
              /clone_lib="Soares NSF F8 9W OT PA_P_S1"
              /notes="Organ: pooled; Vector: p773D-Pac (Pharmacia) with
              a modified polylinker; Site_1: Not 1; Site 2: Eco RI;
              Equal amounts of plasmid DNA from five normalized
              libraries were mixed, and ss circles were made in vitro.
              Following HAP purification, this DNA was used as tracer in
              a subtractive hybridization reaction. The driver was from
              PCR-amplified cDNAs from pools of 5,000 clones made from
              the same 5 libraries. The pools consisted of the following
              libraries and clonesIDs: Soares NBHSF pool 1:
              30384-31019, 323208-325895 Soares NB2HP pool 1:
              145032-147335, 147720-148103, 148872-149255, 15002 -
              150407, 151176-152327 Soares NB2HF8-9W pool 1:
              758280-760583, 772104-774407 Soares NBHPA pool 1:
              304776-306311, 320136-322823, 326280-326663 Soares NBHOT
              pool 1: 723720-726407, 739080-740999 Subtraction by Bento
              Soares and M. Fatima Ronaldo."

ORIGIN

Query Match 20.2%; Score 488; DB 9; Length 488;
Best Local Similarity 100.0%; Pred. No. 1.7e-238;
Matches 488; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1876 AGCGTCTGGTCCCTATGGCGCGAGTCCCTCTGAGGACTCTGACGAGAGTCAAAGGGG 1935
Db       488 AGCGTCTGGTCCCTATGGCGCGAGTCCCTCTGAGGACTCTGACGAGAGTCAAAGGGG 429

QY      1936 CTGGCAAGAGAGTAATCGGATTGTGTACGATTGTGAGCTCCCACTCTCCCGGCGAAGATGCC 1995
Db       428 CTGGCAAGAGAGTAATCGGATTGTGTACGATTGTGAGCTCCCACTCTCCCGGCGAAGATGCC 1995

QY      1996 GAAGATGAGAGGCCACTCCGCACAGGCTTCAAGAACCCATGACCTTAAACGGTGTCTAAT 2055
Db       368 GAAGATGAGAGGCCACTCCGCACAGGCTTCAAGAACCCATGACCTTAAACGGTGTCTAAT 309

QY      2056 AGTGCAGACAGCAGCAGTACCCGAAAGAAAAGCGCCTAGGCGCTGTATGTTGCCAGCTGC 2115

```

FEATURES source

ORIGIN

Query Match	19.3%;	Score 465;	DB 13;	Length 516;
Best local similarity	99.5%;			
Pred. No.	1.1e-226;			

[illegible]

DB
316 CAGTGGATCTTCCTGCGCCGCCGACCTGTACGAGGACTCAAGGGGGTTGGCAA 1943

QV
1884 GTTGCCCTATGGCGCCGAGTCTCTGAGGACTCTGACGAGGACTCAAGGGGGTTGGCAA 1943

Db 456 GGTTGCCCTATGGCGCGAGTCCTCTGAGGACTCTGACGAGGACTCAAGGGCTGGCAA 397

QY 1944 GGAGAA TGGGAT TGGTAC GATTGT GAGTCC CACTCT CCCGGC CCAAGAT GCCGAAGATGA 2003

D5	GGAGAAATGGGATTGGTACGAATGGAGGCTCCCGCCCCCGCAGTTTC
396	GGAGAAATGGGATTGGTACGAATGGAGGCTCCCGCCCCCGCAGTTTC
Ov	2004 GGAGGCCACTCCGACGAGCTTCAAGAAGCCCATGACCCATAA CCGTGCTAATAGTGACA 2063

336 GGAGGCCACTNCGACGAGCTTCAAGAACCCATGACCCCTAACGGTGTAAATAGTCAGA 277

QY 2064 CAGCGACGTGACCCGAAAGAAAACGGCCTAGCGCCTGATGGTGCCAGCTGCCAAGGCCA 2123

D_b
276 CAGGACAGTGAACCCCGAAGAATAACGGCTTAGCGCCTGTATGGTGCCCCGCCTTCGCCCAAGAA

O_r
2124 GCCTGCCCTGCACTCATAAAAATCCCATTGTCTTAAGGCAAACGGTCTTTCTCGAAAACITTCAT 2189

216 GCCTGCCCTGCACTCAGAAAATCCCTTTGGCTAGGCAACGGCTTCTCTGGAAAGTTGAT 157

QY 2184 GCCTGCTCCTTTGCTGTCTCTCCAGAGACAAAATCTTAGAGACCTTCAGCTTAGCAA 224

db 156 GCCTGCTCCTTGTCTCTCCAGGAAGCAAAAATCTTAGAGACCTTCAGGCTTAGCAA 37
2044 CAAATCTCAACGGCTCGACGGAAGAAATGAGTGCACCTGGAGCAGAGAGGGGGCCTCCGA 230

QY 2211 CAGAGGCTCGACGATGAATGAGTGCCTGGAGCAGAGGGGCCCTCCCGA 37

Ddb 96 CAACTGAAAGGCTCGACGATGAATGAGTGCCTGGAGCAGAGGGGCCCTCCCGA 37

QY 2304 GGACGCGAGCGCGAGCCTCAGCCTGGCAGCCCGC 2339

Db 36 GGACCGGACGCGAGCTCAGCTGGCAGCCCCG 1

RESULT 26
BU148710

DEFINITION AGENCOURT_8684470 NIH_MGC_47 Homo sapiens
LOCUS BU148710 308 bp mRNA
IMAGE:6377023
5' mRNA sequence.

ACCESSION	BUI48710
VERSION	BUI48710.1
FORM	GI:2662242

REINWORDS	ESL	Source	ORGANISM	Meaning	Example
			Homo sapiens (human)		
			Homo sapiens		

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

Query Match	19.3%	Score 465;	DB 9;	Length 465;
Best Local Similarity	100.0%	Pred. No 1.1e-226;		
Matches 465;	Conservative	0;	Mismatches 0;	Indels 0;
Gaps	0;			
1908	YGAGGACTCTGACGAGGAGTCAAAAGGGGCTGGCGAAGAGAAATGGGATTGGTACGATTGT	1967		
465	YGAGGACTCTGACGAGGAGTCAAAAGGGGCTGGCGAAGAGAAATGGGATTGGTACGATTGT	406		
1968	YGAGTCCCACTCTCCGGCCAGAGATCCGGAAGATGAGGAGGCCACTCCGCACAGACTTCA	2027		
405	YGAGTCCCACTCTCCGGCCAGAGATCCGGAAGATGAGGAGGCCACTCCGCACAGACTTCA	346		
2028	AGAACCCATGACCCCTAAACGGTGTCTAATAGTGCAGACAGCGACAGTGCACCCGAAAGAAA	2087		
345	AGAACCCATGACCCCTAAACGGTGTCTAATAGTGCAGACAGCGACAGTGCACCCGAAAGAAA	286		

2088 CGGCCTAGGCCTGATGTGTCAGCTGCCAAGGCAGCGCTGCCCTGCACCTCAGAAAAATCC 214
|||
285 CGGCCTAGGCCTGATGTGTCAGCTGCCAAGGCAGCGCTGCCCTGCACCTCAGAAAAATCC 236
|||

225 CTTTGCTAAGGCAACCGGTCTTCTCTGGAAAGTGTATGCTCTGCTCTTTTGCTGTCTCTCC 166

165 AGAAGACAAAATCTTAGAGACCTTCAGGCTTAGCAACAACCTGAAGGCTCGACGGATGA 106

2268 AATGAGTGCACCTGGAGCAGAGAGGGGCCCTCCCGAGGACCGCGACGCCGAGCCTCAGCC 2327

105 AATGAGTGACCTGGACAGAGAGGGGCCCTCCCGAGGACCGCGACGCCGAGCCTCAGCC 46

2528 TGGAGCCCGCCGCGGCATCCCTGGAGGAGCAGATCGGCCGC 1

b

RESIT 25

BQ778360/c
LOCUS

DEFINITION mRNA sequence.
ACCESSION BQ778360

VERSION
BQ778360.1
GI:21986832
EST.
KEYWORDS
normal
normal
normal (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Craniata; Vertebrata; Euteleostomi;
Gastropoda; Mollusca; Bivalvia; Lamellibranchia; Heterodonta

REFERENCE

1 (bases 1 to 516)

Mammalla; Eutneria;
Filinats; Catalinini; Montanaro; Aromat.

AUTHORS

Kentz, G., Permutt, A., Lee, C., Kaestner, K.,
Melton, D., Brown, J.,

Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Gabbler, A., Meisinger, B., Ditter, E., Ronko, J., Bennett, J.,

Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarishvili, R., Williams, T., Jackson, Y. and Bowers, Y.

JOURNAL
TITLE
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

MA 02138
Tel: 617-495-1812

Email: dmelton@biohpc.harvard.edu
Library was constructed by Dr. Hiroshi Inoue
DNA sequencing by:
Fax: 617-495-8557

Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue

(unpublished results),
 Seg primer: -40UP from Gibco
 High quality sequence stop: 439.

REFERENCE	1 (bases 1 to 988)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CM2560 row: c column: 08 High quality sequence stop: 471. Location/Qualifiers
FEATURES	1..988
source	/organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE:6377023" /tissue_type="neuroblastoma, cell line" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_47" /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN	
Query Match	19.1%; Score 460; DB 13; Length 988;
Best Local Similarity	99.6%; Pred. No. 4.7e-224;
Matches 560; Conservative	0; Mismatches 2; Indels 0; Gaps 0
QY	515 GTCATGCAGAAAGCGCTTTGTATGATGTGTACAATGCAGACACATATTACCCAGGCACCTCA 574
Db	33 GTCATGCAGAAAGCGCTTTGTATGATGTGTACAATGCAGACATATTACCCAGGCACCTCA 92
QY	575 GTAATCCTGGGAGCTTATTAAACCAATGTTGTTCATCAATGAGATGCGGGTATAGCTA 634
Db	93 GTAATCCTGGGAGCTTATTAAACCAATGTTGTTCATCAATGAGATGCGGGTATAGCTA 152
QY	635 GGCACCTCCGCTTTTGGAAACCAAGAAATGCCATGAATTCCTTCATATACCTGTTGATG 694
Db	153 GGCACCTCCGCTTTTGGAAACCAAGAAATGCCATGAATTCCTTCATATACCTGTTGATG 212
QY	695 CTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCA 754
Db	213 CTATGCAGAAAGCATGCTTGAATGGCAGCAATAAATTAGACAGACACACCCAGGCCACCA 272
QY	755 CTCCTGTTTGTCCAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTGCA 814
Db	273 CTCCTGTTTGTCCAGATATTTGGAGGATACCTAAGATCTAGAGTCAAAATGTTTAAATTGCA 332
QY	815 AGGGCGTTTCAGATACCTTTGATCCCATATCTTGATATATAACATTGGAGATAAAGCGTCTC 874
Db	333 AGGGCGTTTCAGATACCTTTGATCCCATATCTTGATATATAACATTGGAGATAAAGCGTCTC 392
QY	875 AGAGTGTCAACAGGCAATTTGGACAGTTTCTGTAAGCCGGAACAGCTTGATGGAGAAACT 934
Db	393 AGAGTGTCAACAGGCAATTTGGACAGTTTCTGTAAGCCGGAACAGCTTGATGGAGAAACT 452
QY	935 CGTACAGTGCAGCAAGTGTAATAAGATGGTTCCAGCTTCAAAAGAGGTTTCACTATCCATA 994
Db	453 CGTACAGTGCAGCAAGTGTAATAAGATGGTTCCAGCTTCAAAAGAGGTTTCACTATCCATA 512
QY	995 GATCCTCTTAATGTTCTTATACATTTCTCTGAAACGTTTTTGCAAAATTTTACCGGTGGAAAA 1054

2167 CTTCTGAAAGTTGATCGCTGCTCTTTTGTGTCTCTCCAGAGACAAATCTTAG 2226
 206 CTTCTGAAAGTTGATCGCTGCTCTTTTGTGTCTCTCCAGAGACAAATCTTAG 147
 2227 ACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAATGAGTGCACCTGGAGCA 2286
 146 ACCTTCAGGCTTAGCAACAACTGAAGGCTCGACGGATGAATGAGTGCACCTGGAGCA 87
 2287 GAGAGGGCCCTCCGAGAGCCGACGACGAGCTCTCAGCCCTGGCAGCCCGCCCGGAA 2346
 86 GAGAGGGCCCTCCGAGAGCCGACGACGAGCTCTCAGCCCTGGCAGCCCGCCCGGAA 27
 2347 TCCCTGGAGGAGCCAGATGCGGCCGC 2372
 26 TCCCTGGAGGAGCCAGATGCGGCCGC 1

BI914321 625 bp mRNA linear EST 17-OCT-2001
 603180996F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5245295 5',
 mRNA sequence.
 BI914321 GI:16198830
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM11618 row: o column: 24
 High quality sequence stop: 620.
 Location/Qualifiers
 1..625
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5245295"
 /lab_host="DH10B"
 /clone_lib="NIH_MGC_121"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: EcoRV (destroyed); RNA source anonymous pool of 3
 fetal brains, female age 20 weeks, female age 24 weeks,
 and male age 26 weeks. Library is oligo-dT primed and
 directionally cloned (EcoRV site is destroyed upon
 cloning). Average insert size 1.7 kb, insert size range
 0.7-3.5 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 017. Note:
 this is a NIH_MGC Library."

ORIGIN
 Query Match 18.4%; Score 444; DB 12; Length 625;
 Best Local Similarity 100.0%; Pred. No. 6.9e-216;
 Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

88 AAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGATCAGCTGGCAGCTCCGAGCA 147
 63 AAAGCTTCTGAATCTTCAGACCCATCAGCCTATCAGATCAGCTGGCAGCTCCGAGCA 122

148 GTCTCACCTGGAGACATGGATCGAGTTCTGCCAGCTGGGTCGTGTCTTCAATGAAT 207
 123 GTCTCACCTGGAGACATGGATCGAGTTCTGCCAGCTGGGTCGTGTCTTCAATGAAT 182
 208 GATGTCTCAATCACAACATCTTTTAGGACAGTACCTGGTGTCTAGTTTATTTCGAGT 267
 183 GATGTCTCAATCACAACATCTTTTAGGACAGTACCTGGTGTCTAGTTTATTTCGAGT 242
 268 TCATCTGTACCTGATAAATCAAAACCATCACCACAAAAGGATCAAGCCCTAGGTGATGC 327
 243 TCATCTGTACCTGATAAATCAAAACCATCACCACAAAAGGATCAAGCCCTAGGTGATGC 302
 328 ATCGCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAACA 387
 303 ATCGCTCTCTCCACAGAAAGTTCTTTTCCCATCTGAGAAGATTGTCTTAAGTGGCAACA 362
 388 ACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATGCAGCA 447
 363 ACTCATAGATTGGAGCTGGGCTCCAGAAATTTGGGCAATACCTGTTTCCCAATGCAGCA 422
 448 CTGCACTGTTTAACTTACCTACACACACCTCTTGCCTTATCATGCTATCATGAACACTCC 507
 423 CTGCACTGTTTAACTTACCTACACACACCTCTTGCCTTATCATGCTATCATGAACACTCC 482
 508 AAAACATGTCATGCAGAGGCTTT 531
 483 AAAACATGTCATGCAGAGGCTTT 506

RESULT 29
 BQ956351
 LOCUS
 DEFINITION
 5', mRNA sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1CM2617 row: g column: 19
 High quality sequence stop: 473.
 Location/Qualifiers
 1..1085
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6429354"
 /tissue_type="epidermoid carcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_101"
 /note="Organ: lung; Vector: pOTB7; Site 1: EcoRI; Site 2:
 XhoI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCACGAG(G). Library constructed by Ling Hong in the
 laboratory of Gerald M. Rubin (University of California,
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
 Superscript II RT (Life Technologies). Note: this is a
 NIH_MGC Library."

FEATURES
 source

ORIGIN

	Query Match	18.1%; Score 436; DB 13; Length 1085;	
	Best Local Similarity 99.6%; Pred. No. 1e-211;		
	Matches 536; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	431	GTTTGGCAATGCAGCACCTGACGTGTTTAACTTCACACACCACCTCTTGCCAAATTACATGC	490
Db	11		
QY	491	TATCACAATGAACACTCCAACCATGTGTCAGAGAAGCGTTTTGTATGATGTGTACAAATGC	550
Db	71		
QY	551	RAGCACATAATCCCAGGCACTCAGTAATCTCTGGGACGTTATTAACCAATGTTTGTC	610
Db	131		
QY	611	TCAATCAGATGGGGCGTAGCTTAGGCACCTCCGTTTTGGAACCAAGAGATGCCCATG	670
Db	191		
QY	671	AATTCTTCAAATACACTGTGTGATGCTATGCAGAAAAGCATGTTGAATGGCAGCAATAAT	730
Db	251		
QY	731	TAGACAGACACACCCAGGCCACCACTCTGTTGTGCAGATATTTGGAGGATACCTAAGAT	790
Db	311		
QY	791	CTAGAGTCAAAATGTTTAAATTCGAAGGGCGTTTCAGATACTTTTGATCCATATCTTGATA	850
Db	371		
QY	851	TAA CATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGACGAGTTTGTGAAGC	910
Db	431		
QY	911	CGGAACAGCTTGTNGAGAAAACTCGTACAGTGCAGCAAGTGTAAAAAGATGGTTCC	968
Db	491		
QY		CGCAACAGCTTGATGGAGAAACTCGTACAGTGCAGCAAGTGTAAAAAGATGGTTCC	548

```

RESULT 30
LOCUS      B1914337
DEFINITION 603182326F1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5246166 5',
            B1914337
ACCESSION  B1914337
VERSION    B1914337.1 GI:16178569
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 719)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgabs-remail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11621 row: d column: 07
            High quality sequence stop: 535.
            Location/Qualifiers
                1..719
                /organism="Homo sapiens"
FEATURES
source

```

FEATURES
source

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5246166"
/lab_host="DH10B"
/clone_lib="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH MGC library."

```

ORIGIN

Query Match	18.0%;	Score 433;	DB 12;	Length 719;
Best Local Similarity	100.0%;	Pred. No. 3.2e-210;		
Matches 433;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GGGCTGTGTGCGGGCGGGCGGGCGGGGATGGAGCGAGCGCGGACCGCGGT	60	
Db	26	GGGCTGTGTGCGGGCGGGCGGGCGGGGATGGAGCGAGCGCGGACCGCGGT	85	
QY	61	CAGAGTTGAACAAATGACCAATGATTGACAAAGCTTCTGAATCTTTCAGACCCATCAGCCGTAT	120	
Db	86	CAGAGTTGAACAAATGACCAATGATTGACAAAGCTTCTGAATCTTTCAGACCCATCAGCCGTAT	145	
QY	121	CAGAATCAGCTGGGAGCTCCGAGGCGAGTCTACCTGGAGACATGATGCAGGTTCTGGCC	180	
Db	146	CAGAATCAGCTGGGAGCTCCGAGGCGAGTCTACCTGGAGACATGATGCAGGTTCTGGCC	205	
QY	181	AGCTGGGGTGCTGTGTCTTTCATTGAATGATGTGTCAAATCACAACATCTTCTTTAGGACCA	240	
Db	206	AGCTGGGGTGCTGTGTCTTTCATTGAATGATGTGTCAAATCACAACATCTTCTTTAGGACCA	265	
QY	241	GTACCTGTGTGTGTAGTTTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA	300	
Db	266	GTACCTGTGTGTGTAGTTTATTCGAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA	325	
QY	301	CAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTTCCCATCT	360	
Db	326	CAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCTCCACAGAAAGTTCTTTTCCCATCT	385	
QY	361	GAGAAGATTGTCTTAAGTGGCAACAACTCATAGTTGGAGCTGGGCTCCAGAATTTC	420	
Db	386	GAGAAGATTGTCTTAAGTGGCAACAACTCATAGATTGGAGCTGGGCTCCAGAATTTC	445	
QY	421	GGCAATACCTGTT	433	
Db	446	GGCAATACCTGTT	458	

RESULT 31
AI418577/c

LOCUS	AT1418577	466 bp	mRNA	linear	EST 30-MAR-1999
DEFINITION	t69n12.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:2104583 3', mRNA sequence.				
ACCESSION	AI418577				
VERSION	AI418577.1	GI:4264508			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	1 (bases 1 to 466)				
AUTHORS	NCI/NINDS-CGAP	http://www.ncbi.nlm.nih.gov/ncicgap.			
TITLE	National Cancer Institute / National Institute of Neurological Disorders and Stroke, Brain Tumor Genome Anatomy Project (CGAP/BTCAP), Tumor Gene Index				
JOURNAL	Unpublished (1998)				
COMMENT	Contact: Robert Strausberg	Ph D			

Email: cgapbs-remail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 858 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 453.
Location/Qualifiers

FEATURES

source

1. 466
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2104583"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI CGAP Brn23"
/note="Organ: brain; Vector: pTVT3D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTCAAGTGGGAGCGCGCATATCTTTTCTTTTCTTTTCTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pTVT3 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 17.4%; Score 420; DB 9; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.3e-203;
Matches 420; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1925 AGTCAAAGGGGCTGGCAGGAGAAATGGGATTGGTACGATTGTGAGCTCCACTCTCCCG 1984
DB 420 AGTCAAAGGGGCTGGCAGGAGAAATGGGATTGGTACGATTGTGAGCTCCACTCTCCCG 361
QY 1985 GCCAAGATCCGAGATGAGAGGCGCACTCCGACGAGCTTCAAGACCCATGACCCCTAA 2044
DB 360 GCCAAGATCCGAGATGAGAGGCGCACTCCGACGAGCTTCAAGACCCATGACCCCTAA 301
QY 2045 ACGGTGCTAATAGTGACAGACGACAGTACCGAAGAAACCGGCTAGCGCTGATG 2104
DB 300 ACGGTGCTAATAGTGACAGACGACAGTACCGAAGAAACCGGCTAGCGCTGATG 241
QY 2105 GTGCCAGCTGCAAGGCGAGCTGCCCTGCACTCAGAAATCCCTTTGCTAAGGCAACG 2164
DB 240 GTGCCAGCTGCAAGGCGAGCTGCCCTGCACTCAGAAATCCCTTTGCTAAGGCAACG 181
QY 2165 GTCTTCTGGAAGTTGATGCTCTCTCTTTGCTGCTCTCCAGAGACAAATCTTTAG 2224
DB 180 GTCTTCTGGAAGTTGATGCTCTCTCTTTGCTGCTCTCTCCAGAGACAAATCTTTAG 121
QY 2225 AGACCTTCAGGCTTAGCAACAACTGAAAGGCTCGACGATGAATGAGTGCACCTGGAG 2284
DB 120 AGACCTTCAGGCTTAGCAACAACTGAAAGGCTCGACGATGAATGAGTGCACCTGGAG 61
QY 2285 CAGAGAGGGGCTCCCGAGACCGGACGCGGAGCTCAGCTGGCAGCCCGCCGCG 2344
DB 60 CAGAGAGGGGCTCCCGAGACCGGACGCGGAGCTCAGCTGGCAGCCCGCCGCG 1

RESULT 32
AW024899/c 552 bp mRNA linear RST 27-OCT-1999
LOCUS wu92d12.x1 NCI CGAP_kid3 Homo sapiens cDNA clone IMAGE:2527511 3',
DEFINITION mRNA sequence.
ACCESSION AW024899
VERSION AW024899.1 GI:5878429

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 552)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE

Tumor Gene Index

JOURNAL

Unpublished (1997)

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 490.
Location/Qualifiers

1. 552
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2527511"
/lab_host="DH10B"
/clone_lib="NCI CGAP Kid3"
/note="Organ: kidney; Vector: pTVT3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pTVT3 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

ORIGIN

Query Match 17.4%; Score 419; DB 9; Length 552;
Best Local Similarity 99.8%; Pred. No. 4.6e-203;
Matches 539; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1834 TGCTCCAGCCCGTATGAATGCAATCCAAATCCAGCTGAACTCCAGCGTCTGGTCCCTAT 1893
DB 540 TGCTCCAGCCCGTATGAATGCAATCCAAATCCAGCTGAACTCCAGCGTCTGGTCCCTAT 481
QY 1894 GGGCCCGAGTCTCTGAGGACTCTGACGAGAGTCAAAGGGGCTGGGCAAGGAGATGGG 1953
DB 480 GGGCCCGAGTCTCTGAGGACTCTGACGAGAGTCAAAGGGGCTGGGCAAGGAGATGGG 421
QY 1954 ATTGGTACGATTGTGAGCTCCACTCTCCGCGCAAGATCCGAGATGAGAGGCCACT 2013
DB 420 ATTGGTACGATTGTGAGCTCCACTCTCCGCGCGCAAGATCCGAGATGAGAGGCCACT 361
QY 2014 CCGCAGCAGCTTCAAGAACCCATGACCTTAAACCGTGTCTATAGTGACAGACGACAGT 2073
DB 360 CCGCAGCAGCTTCAAGAACCCATGACCTTAAACCGTGTCTATAGTGACAGACGACAGT 301
QY 2074 GACCCGAAAGAAAACGGCCTAGCCGC-TGATGTTCCAGCTGCCAAGCCAGCTGCCCT 2132
DB 300 GACCCGAAAGAAAACGGCCTAGCCGCCTGATGTTGCCAGCTGCCAAGCCAGCTGCCCT 241
QY 2133 GCACCTCAGAAATCCCTTTGCTAGGCAACCGTCTTCTCGAAAGTTGATGCTGCTCC 2192
DB 240 GCACCTCAGAAATCCCTTTGCTAGGCAACCGTCTTCTCGAAAGTTGATGCTGCTCC 181
QY 2193 TTTGCTGCTCTCCAGAGACAAATCTTATAGACCTTTCAGGCTTAGCAACAACTGAA 2252
DB 180 TTTGCTGCTCTCCAGAGACAAATCTTATAGACCTTTCAGGCTTAGCAACAACTGAA 121

```

QY 2253 AGGCTCAGCGATGAAATGAGTGCACCTGGAGCAGAGGGGCCCTCCGAGGACCGGA 2312
Db 120 AGGCTCAGCGATGAAATGAGTGCACCTGGAGCAGAGGGGCCCTCCGAGGACCGGA 61
QY 2313 CGCGAGCCTCAGCTGGCAGCCCCCGCGCGGATCCTCGAGGAGCAGATCGGCGCG 2372
Db 60 CGCGAGCCTCAGCTGGCAGCCCCCGCGCGGATCCTCGAGGAGCAGATCGGCGCGC 1

RESULT 33
AG116030/c
LOCUS AG116030 657 bp DNA linear GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-123H03.R, genomic survey sequence.
ACCESSION AG116030
VERSION AG116030.1 GI:16736549
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 657)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimpansegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/;
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI.
Location/Qualifiers
1..657
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-123H03.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
Query Match 17.2%; Score 415; DB 29; Length 657;
Best Local Similarity 99.5%; Pred. No. 5.4e-201;
Matches 565; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1457 AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACACGCCAAGCGATTCATGT 1516
Db 621 AGAATCCACCTCACTTAATGGGACTGGACCATTTGAAGACACGCCAAGCGATTCATGT 562
QY 1517 CGAGTCTTAACGGGAATTCACAGTGTCAACAGGGCTAGTCTCTTTAATGCTTCAGCTTCTG 1576
Db 561 CGAGTCTTAACGGGAATTCACAGTGTCAACAGGGCTAGTCTCTTTAATGCTTCAGCTTCTG 502
QY 1577 TCCAAACTGGTCTAGTTAATAGGTCCTCAGTGATPCCAGAACATCTTAAGAAAAA 1636
Db 501 TCCAAACTGGTCTAGTTAATAGGTCCTCAGTGATPCCAGAACATCTTAAGAAAAA 442
QY 1637 TTCAATCAGTANTCAACAAAGTTGCTGTTGCCAGTGTCACTCAACCTTAACCTTC 1696
Db 441 TTCAATCAGTANTCAACAAAGTTGCTGTTGCCAGTGTCACTCAACCTTAACCTTC 382

```

```

QY 1697 ATAGTAATCTTTGGAGAACCTTACCAAGCCCGTTCCCTCTTTACCAATACCAATTCG 1756
Db 381 ATAGCAATCTTTGGAGAACCTTACCAAGCCCGTTCCCTCTTTACCAATACCAATTCG 322
QY 1757 CAGTACAGTCTACCTCGAAGCGATCTACGATGTCTAGTAAAGTAAACAAACCGA 1816
Db 321 CAGTACAGTCTACCTCGAAGCGATCTACGATGTCTAGTAAAGTAAACAAACCGA 262
QY 1817 TCCCCCGCAGTGAATCTGCTCCAGCCCGTGAATGCAATCCAAATCCAAATCCAAATCC 1876
Db 261 TCCCCCGCAGTGAATCTGCTCCAGCCCGTGAATGCAATCCAAATCCAAATCCAAATCC 202
QY 1877 GCCTGCTGTTGCCCTATGGCGCCGAGTCTCTGAGGACTCTGAGGAGTCTCAAGGGGC 1936
Db 201 GCCTGCTGTTGCCCTATGGCGCCGAGTCTCTGAGGACTCTGAGGAGTCTCAAGGGGC 142
QY 1937 TGGGCAAGAGAAATGGATTTGGTACGATTTGAGTCTCCCACTCTCCGCGCAAGATGCCG 1996
Db 141 TGGGCAAGAGAAATGGATTTGGTACGATTTGAGTCTCCCACTCTCCGCGCAAGATGCCG 82
QY 1997 AAGATGAGGAGGCCACTCCGCAAGAGCT 2024
Db 81 AAGATGAGGAGGCCACTCCGCAAGAGCT 54

RESULT 34
AG121934
LOCUS AG121934 MAMMAL Homo sapiens cDNA clone MAMMAL001308 5', mRNA
DEFINITION sequence.
ACCESSION AG121934
VERSION AG121934.1 GI:10937169
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saiko, K., Kawai, Y.,
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
Isogai, T.
HRI human cDNA project
Unpublished (2000)
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MAMMAL001308"
/tissue_type="mammary gland"
/clone_lib="MAMMAL"
/vector="Vector: pME18SFL3"

ORIGIN
Query Match 16.9%; Score 407; DB 9; Length 818;
Best Local Similarity 99.7%; Pred. No. 7.2e-197;
Matches 577; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 798 CAAATGTTTAAATGCAAGGGCGTTTCAGATATCTTTGATCCATATCTTGATATAACATT 857
Db 72 CAAATGTTTAAATGCAAGGGCGTTTCAGATATCTTTGATCCATATCTTGATATAACATT 131

```



```

Seq primer: M13 Forward
POLYA=No.

FEATURES
    source
        Location/Qualifiers
            1..379
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:3070478"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="NCI CGAP Sub7"
                /note="Vector: pTVT3D-Pac (Pharmacia) with a modified
                polylinker. Site 1: Not I; Site 2: Eco RI; NCI CGAP Sub7
                is a subtracted library derived from NCI CGAP Sub6. The
                NCI CGAP Sub7 library had 12 million recombinants. A
                single-stranded DNA preparation of NCI CGAP Sub6 was used
                as a tracer in a subtractive hybridization with a driver
                comprising: the IMAGE pool (NCI CGAP Kid3 pool 1 LLAM
                3334-3337, 3682-3683, 3798-3803) (IMAGE CloneIDs
                1323176-13231911, 1456008-1456775, 1500552-1502855);
                NCI CGAP Kid5 pool 1 LLAM 3338-3342, 3722-3725, 3776-3778
                (IMAGE CloneIDs 1323912-1325831, 1471368-1472903,
                1492104-1493255); NCI CGAP Lu5 pool 1 LLAM 3575-3582,
                3851-3854 (IMAGE CloneIDs 1414920-1417991,
                1520904-1522439); NCI CGAP GC4 pool 1 LLAM 3164-3167,
                3716-3720, 3733-3735 (IMAGE CloneIDs
                1257096-1258631, 1469064-1470983, 1475592-1476743);
                NCI CGAP Pr22 pool 1 LLAM 2457-2459, 2758-2759, 3062-3068
                (IMAGE CloneIDs 985608-986759, 1101192-1101959,
                1217928-1220615); NCI CGAP_Colo pool 1 LLAM 2644-2653,
                2871-2872 (IMAGE CloneIDs
                1057416-1061255, 1144584-1144535). (6% of the driver
                population), plus a pool of 3 840 arrayed clones from
                NCI CGAP Sub1 (IMAGE CloneIDs 2708616-2710535) and
                NCI CGAP Sub2 (IMAGE CloneIDs 2710536-2712455) (4% of
                the driver population), plus a pool of 11,136 clones from
                NCI CGAP Sub3 (IMAGE CloneIDs 2712456-2723591) (10% of
                the driver population), plus a pool of 5,472 clones from
                NCI CGAP Sub4 (IMAGE CloneIDs 2723592-2729326) (40% of the
                driver population), plus a pool of 4032 clones from
                NCI CGAP Sub6 (IMAGE CloneIDs 2728969-2733190) (40% of the
                driver population). Subtraction was performed as
                previously described [Bonaldo, Lennon & Soares (1996) :
                Normalization and Subtraction: Two Approaches To
                Facilitate Gene Discovery. Genome Research 6, 791-806.
                TAG SEQ=None found"

```

Query Match	15.3%	Score 370	DB 10	Length 379	Best Local Similarity 100.0%	Pred. No. 5.9e-178	Mismatches 0	Indels 0	Gaps 0
Matches 370	Conservative								
QY	2002	GAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCCCTAAACCGTGCTAATAGTSCA	2061						
Db	370	GAGGAGGCCACTCCGACAGAGCTTCAAGAACCCATGACCCCTAAACCGTGCTAATAGTSCA	311						
QY	2062	GACAGCGACAGTGACCCGAGAGAAAACGCCCTAGCGCCTGATGTGCCAGTGCCTAAGGC	2121						
Db	310	GACAGCGACAGTGACCCGAGAGAAAACGCCCTAGCGCCTGATGTGCCAGTGCCTAAGGC	251						
QY	2122	CAGCCTGCCCTGCACCTCAGAAAAATCCCTTTCCTAAGCGAAACGGCTCTTCTCGGAAAGTTG	2181						
Db	250	CAGCCTGCCCTGCACCTCAGAAAAATCCCTTTCCTAAGCGAAACGGCTCTTCTCGGAAAGTTG	191						
QY	2182	ATGCGCTGCTCTCTTTTCTCTCTCTCCAGAGACAAAAATCTTTAGAGACCTTCAGCGCTTAGC	2241						
Db	190	ATGCGCTGCTCTCTTTTCTCTCTCTCCAGAGACAAAAATCTTTAGAGACCTTCAGCGCTTAGC	131						
QY	2242	AACAACTGAAAGGCTCGACGGATGAATGATGACACTCGAGACAGAGAGGGGCGCTCCC	2301						
Db	130	AACAACTGAAAGGCTCGACGGATGAATGATGACACTCGAGACAGAGAGGGGCGCTCCC	71						
QY	2302	GAGGACCGCGAGCGCGAGCCTCAGCCTGGACGCCCGCCCGAATCCCTTGGAGGAGCCA	2361						

Dn	273	TGATGTGCCAGCTCCCAAGCGCAGCTGCCTGCCCTGCACCTCAGAAAATCCCTTGTGAAGGC	214
Qy	2160	AAACGGTTCTCTGGAAGTTGATGCCTTCCTCTTGCTGTCTCTCCAGAACAAAAAT	2219
Dn	213	AAACGGTTCTCTGGAAGTTGATGCCTTCCTCTTGCTGTCTCTCCAGAACAAAAAT	154
Qy	2220	CTTAGAGACTTTCAGGCTTTAGCAACAATAAAGGCTCGAGCGATGAAATGAGTGCACC	2279
Dn	153	CTTAGAGACTTTCAGGCTTTAGCAACAATAAAGGCTCGAGCGATGAAATGAGTGCACC	94
Qy	2280	TGGAGCAGAGAGGGGCTCCCGAGGACCGGACCGGAGCTTCAGCCTGGCAGCCC	2336
Dn	93	TGGAGCAGAGAGGGGCTCCCGAGGACCGGACCGGAGCTTCAGCCTGGCAGCCC	37
RESULT	39		
LOCUS	AA255809		
DEFINITION	zr3h12_r1 Soares NhhMPu_S1 Homo sapiens cDNA clone IMAGE:682343	linear	EST 06-AUG-1997
ACCESSION	AA255809	5', mRNA sequence.	
VERSION	AA255809.1	GI:1891350	
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	1 (bases 1 to 416)		
	Hillier,L., Allen,M., Le,N., Lennon,G., Marra,M., Martin,J.,		
	Kucaba,T., Lacy,M., Li,S., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,		
	Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,		
	White,Y., Wyllie,T., Waterston,R. and Wilson,R.		
TITLE	WashU-Merck EST Project 1997		
JOURNAL	Unpublished (1997)		
COMMENT	Contact: Wilton RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LLNL ; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert Length: 849 Std Error: 0.00		
	Seq primer: -28ml3 rev2.ET from Amersham.		
FEATURES	Location/Qualifiers		
source	1..416		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:682343"		
	/tissue_type="Pooled human melanocyte, fetal heart, and		
	pregnant uterus"		
	/lab_host="DH10B"		
	/clone_lib="Soares NhhMPu S1"		
	/note="Organ: mixed (see below); Vector: pTZ19-Pac		
	(Pharmacia) with a modified polylinker; Site 1: Not I;		
	Site 2: Eco RI; Equal amounts of plasmid DNA from three		
	normalized libraries (melanocyte 2NDHM, pregnant uterus		
	NhhPU, and fetal heart Nhh19W) were mixed, and ss circles		
	were made in vitro. Following HAP purification, this DNA		
	was used as tracer in a subtractive hybridization		
	reaction. The driver was PCR-amplified cDNAs from pools of		
	5,000 clones made from the same 3 libraries. The pools		
	consisted of I.M.A.G.E. clones 260232-265223,		
	340488-345479, and 484488-489479."		
ORIGIN			
Query Match	14.8%; Score 358; DB 9; Length 416;		
Best Local Similarity	100.0%; Pred.No. 8.7e-172;		
Matches 358; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
QY	1538	GTGTCAACAGGGTAGTCTCTTAATGCTTCAGCTTCTGTCCAAAACCTGTCAGTTAATA	1597

```

Db 1 GTGTCAACAGGGCTAGTCCTGTTAATGCTTCAGCTTCTGTCCAAATAGTGTCAAGTAA 60
QY 1598 GGTCTCAGTATCCAGAGATCCTTAAGAAACAAAATACATCAGTATTCACAAACA 1657
Db 61 GGTCTCAGTATCCAGAGATCCTTAAGAAACAAAATACATCAGTATTCACAAACA 120
QY 1658 AGTGTCTGTTCGCCAGTGTGAGTCTCAACCTCAACCTTCACTAGTAATCTTTGAGAAC 1717
Db 121 AGTGTCTGTTCGCCAGTGTGAGTCTCAACCTCAACCTTCACTAGTAATCTTTGAGAAC 180
QY 1718 CTACCAAGCCGCTCTCTTCTACCATTAACCAATTCAGTATTCAGTATTCAGTATTC 1777
Db 181 CTACCAAGCCGCTCTCTTCTACCATTAACCAATTCAGTATTCAGTATTCAGTATTC 240
QY 1778 CATCTACGATGTGAGTCTCTAGTAAAGTACAAACCGATCCCGCCGAGTGAATCTGTCT 1837
Db 241 CATCTACGATGTGAGTCTCTAGTAAAGTACAAACCGATCCCGCCGAGTGAATCTGTCT 300
QY 1838 CCAGCCCGTGTGAATGGCAATCAACCTCAACCTGAGTGTGCTGGTCCCTATGG 1895
Db 301 CCAGCCCGTGTGAATGGCAATCAACCTGAGTGTGCTGGTCCCTATGG 358

RESULT 40
AA054177
LOCUS
DEFINITION
  zf51b04.r1 Soares retina N2b4HR Homo sapiens cDNA clone
ACCESSION
  AA054177
VERSION
  AA054177.1 GI:1545100
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 440)
  Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
  Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marr, M.,
  Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
  Trevaaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and
  Wilson, R.
  The WashU-Merck EST Project
  Unpublished (1995)
  Contact: Wilson RK
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: est@watson.wustl.edu
  This clone is available royalty-free through LLNL; contact the
  IMAGE Consortium (info@image.llnl.gov) for further information.
  Seq primer: -28M13 rev2 from Amersham.
  Location/Qualifiers
    1..440
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="GDB:1288688"
      /db_xref="taxon:9606"
      /clone="IMAGE:380431"
      /sex="male"
      /tissue_type="retina"
      /dev_stage="55 year old"
      /lab_host="DH10B (ampicillin resistant)"
      /clone_lib="Soares retina N2b4HR"
      /note="Organ: eye; Vector: p7T3D (Pharmacia) with a
      modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
      strand cDNA was primed with a Not I - oligo(dT) primer [5'
      TGTTACCAATCTCAAGTGGAGCGCGCGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT
      double-stranded cDNA was size selected, ligated to Eco RI
      adapters (Pharmacia), digested with Not I and cloned into
      the Not I and Eco RI sites of a modified p7T3 vector
      (Pharmacia). The retinas were obtained from a 55 year old

```

ORIGIN

```

Query Match      14.1%; Score 340; DB 9; Length 440;
Best Local Similarity 100.0%; Pred. No. 1.5e-162;
Matches 340; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 CATCAGCCTATCAGATCAGCTGGCAGCTCCGAGCAGTCTCACCTGGAGACATGATG 169
Db 101 CATCAGCCTATCAGATCAGCTGGCAGCTCCGAGCAGTCTCACCTGGAGACATGATG 160
QY 170 CAGGTTCTGCAGCTGGGCTGCTCTTCATTGAATGATGTCAATCACAACATTT 229
Db 161 CAGGTTCTGCAGCTGGGCTGCTCTTCATTGAATGATGTCAATCACAACATTT 220
QY 230 CTTTAGGACCACTACCTGGTGTGATGTTTATTCGAGTTTCATCTGTACCTGATAAATCAA 289
Db 221 CTTTAGGACCACTACCTGGTGTGATGTTTATTCGAGTTTCATCTGTACCTGATAAATCAA 280
QY 290 AACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCACAGAAAGTTC 349
Db 281 AACCATCACCAAAAGGATCAAGCCCTAGGTGATGGCATCGCTCCACAGAAAGTTC 340
QY 350 TTTTCCCATCTCAGAGATTTCTCTTAGTGGCAACAACTCATAGTTGAGCTGGGC 409
Db 341 TTTTCCCATCTCAGAGATTTCTCTTAGTGGCAACAACTCATAGTTGAGCTGGGC 400
QY 410 TCCAGAAATTTGGCAATATACCTGTTTGGCAATGCAGCACT 449
Db 401 TCCAGAAATTTGGCAATATACCTGTTTGGCAATGCAGCACT 440

RESULT 41
CB990121
LOCUS
DEFINITION
  IMAGE:3033487 5', mRNA sequence.
ACCESSION
  CB990121
VERSION
  CB990121.1 GI:30284641
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 861)
  NIH-MGC http://mgi.nci.nih.gov/.
  National Institutes of Health, Mammalian Gene Collection (MGC)
  Unpublished (1999)
  Contact: Robert Strausberg, Ph.D.
  Email: cgapbs@mail.nih.gov
  Tissue Procurement: Dr. Stefan Hanson
  cDNA Library Preparation: Michael J. Brownstein (NHGRI) with help
  and advice from Piero Carninci (RIKEN)
  cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
  DNA Sequencing by: Agencourt Bioscience Corporation
  Clone distribution: MGC clone distribution information can be
  found through the I.M.A.G.E. Consortium/LLNL at:
  http://image.llnl.gov
  Plate: NDAM378 row: d column: 24
  High quality sequence stop: 339.
  Location/Qualifiers
    1..861
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:30343487"
      /tissue_type="Human Placenta"
      /lab_host="DH10B Tona"
      /clone_lib="NIH_MGC_147"

```

Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."

/notes="Organ: placenta; Vector: pluscriptPR; Site: 1: all-XhoI; Site:2: BamH; Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTT-3', size-selected for average insert size 2.3 kb and normalised to ROT 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIMH/NHGRI, National Institutes of Health). Note: This is a NIH MGC library."

ORIGIN	Query Match	14.1%;	Score 340;	DB 14;	Length 861;
	Best Local Similarity	99.7%;	Pred. No. 1.8e-162;		
	Matches 390;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	63	GAGTGTGAACAATGACCATAGTTGCACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCA	122		
Db	80	GAGTTGAACAANTGACCATAGTTGCACAAAGCTTCTGAATCTTCAGACCCATCAGCCTATCA	139		
QY	123	GAATCAGCCTGGGAGCTCCGAGGCAGTCTCACCTGGAGAGATGATCAGGTTCTGCGAG	182		
Db	140	GAATCAGCCTGGGAGCTCCGAGGCAGTCTCACCTGGAGAGATGATGAGGTTCTGCGAG	199		
QY	183	CTGGGGTGCTGTGTCCTTCAATTGAATGATGTCAAATCACACACTTTTCTTTPAGCACGAT	242		
Db	200	CTGGGGTGCTGTGTCCTTCAATTGAATGATGTCAAATCACACACTTTTCTTTPAGCACGAT	259		
QY	243	ACCTGTGTGCTGTAGTTTATTTCGAGTTCATCTGTACTGTATAAATCAAAACCATCACCACA	302		
Db	260	ACCTGTGTGCTGTAGTTTATTTCGAGTTCATCTGTACTGTATAAATCAAAACCATCACCACA	319		
QY	303	AAAGGATCAAGCCCTAGTGATGGCATCGCTCCTCCACAGAAAGTCTTTTCCCATCTGA	362		
Db	320	AAAGGATCAAGCCCTAGTGATGGCATCGCTCCTCCACAGAAAGTCTTTTCCCATCTGA	379		
QY	363	GAAGATTTGCTTAAAGTGGCAACAAACCTCATAGAGTTGGAGCTGGGCTCCAGAAATTTGGG	422		
Db	380	GAAGATTTGCTTAAAGTGGCAACAAACCTCATAAAGTTGGAGCTGGGCTCCAGAAATTTGGG	439		
QY	423	CAATACCTGTTTTCGAATGAGCACTGCGAG	453		
Db	440	CAATACCTGTTTTCGAATGAGCACTGCGAG	470		

[illegible]

CDNA Clome CS0DC002DD10QPL; cluster=4596.f

ACCESSION BX331803
VERSION GI:30339041
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1144)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 4596.f
For more information about this cluster, see
<http://www.genoscope.cns.fr/>
<http://cluster.cgi?seq=CS0DC002DD10QPL&cluster=4596.f>. Contact :
Feng Liang Email : flliang@lifetech.com URL :
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DC002DD10QPL.
Location/Qualifiers

CS0DC002DD10QPL

FEATURES

```

1. 1144
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DC002H20"
/tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
/clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

[illegible]

RESULT 43	AA016255/c	499 bp	mRNA	linear	EST 29-NOV-1996
LOCUS	ze33c07.s1 Soares retina N2b4HR				
DEFINITION	IMAGE:360780 3', mRNA sequence.				
ACCESSION	AA016255				
VERSION	AA016255.1	GI:1477302			

EST.	EST.
SOURCE	human)
ORGANISM	Homo sapiens
	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
	1 (bases 1 to 499)
REFERENCE	Hillier, L., Clark, N., Dubuque, T., Eliston, K., Hawkins, M.,
AUTHORS	Hollman, M., Hultman, M., Kucaba, T., Le, M., Lemon, G., Marra, M.,
	Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
	Trevaskis, B., Waterston, R., Williamson, A., Wohlmann, P. and
	Wilson, R.

TITLE	WILSON, K.
JOURNAL	The Washu-Merck EST Project
COMMENT	Unpublished (1995) Contact: Wilson RK Washington University School 4444 Forest Park Parkway, B

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel:	314 286 1800
Fax:	314 286 1810
Email:	est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.	
Possible reversed clone: polyT not found	
Insert Length: 1460 Std Error: 0.00	
Seq primer: -40M13 fwd. from AmerSham	
High quality sequence stop: 347.	
Location/Qualifiers	
1..499	
/organism="Homo sapiens"	
/mol_type="mRNA"	
/db_xref="GDB:1277325"	
/db_xref="taxon:9606"	
/clone="IMAGE:360780"	
/sex="male"	
/tissue type="retina"	
/dev stage="55 year old"	
/lab_host="DH10B (ampicillin resistant)"	
/clone_lib="Soares retina N2b4HR"	
/note="Organ: eye; Vector: pTV73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'- TGTTACCAATCTGAAGTGGGAGCGCGCCTTTTTTTTTTTTTTTT 3'] , double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTV73 vector (Pharmacia). The retinas were obtained from a 55 year old Caucasian and total cellular poly(A)+ RNA was extracted 6 hrs after their removal. The retina RNA was kindly provided by Roderick R. McInnes M.D. Ph.D. from the University of Toronto. Library constructed by Bento Soares and M.Fatima Bonaldo."	
ORIGIN	
Query Match 13.1%; Score 317; DB 9; Length 499;	
Best Local Similarity 100.0%; Pred. No. 1e-150;	
Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	2052 TAATAGTGCAGACGACAGTAGTCACC CGAAGAAAACGGCTTAGCCGTGATGGTGCCAG 2111
Db	321 TAATAGTGCAGACGACGACAGTAGTCACC CGAAGAAAACGGCTTAGCCGTGATGGTGCCAG 262
QY	2112 CTGCCAAGGCCAGCCTGCCTTCCTGCACTCAGAAAAATCCCTTTGCTTAAGGCAAACGGTCTTCC 2171
Db	261 CTGCCAAGGCCAGCCTGCCTTCCTGCACTCAGAAAAATCCCTTTGCTTAAGGCAAACGGTCTTCC 202
QY	2172 TGGAAAGTTGATCGCTTGCTCTTTTGTGTCTCTCCAGAGAACAATACTTAGAGACCTT 2231
Db	201 TGGAAAGTTGATCGCTTGCTCTTTTGTGTCTCTCCAGAGAACAATACTTAGAGACCTT 142
QY	2232 CAGGCTTAGCAACAAACTGAAGGCTCGACGATGAATAGTGCACCTGGAGCAGAGAG 2291
Db	141 CAGGCTTAGCAACAAACTGAAGGCTCGACGATGAATAGTGCACCTGGAGCAGAGAG 82
QY	2292 GGSCCTTCGAGGACCGCGACGCGAGCCTCAGCCTGGCAGCCGCGCGGCGGCAATCCCT 2351
Db	81 GGSCCTTCGAGGACCGCGACGCGAGCCTCAGCCTGGCAGCCGCGCGGCGGCAATCCCT 22
QY	2352 GGAGGAGCCAGATGCGG 2368
Db	21 GGAGGAGCCAGATGCGG 5
RESULT 44	
AA456513	
LOCUS	
DEFINITION	
aai3d07.r1 Soares NHMPu_S1 Homo sapiens cdna clone IMAGE:813133	
5', mRNA sequence.	
ACCESSION	
VERSION	
AA456513.1 GI:2179089	
KEYWORDS	
EST.	


```

Db      318 AGGTGATGGCATCCTCTCCACAGAAAGTTCTTTTCCCATCTCAGAAAGATTGTCTTAA 377
QY      378 GTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAA 416
Db      378 GTGGCAACAAACTCATAGAGTTGGAGCTGGGCTCCAGAA 416

RESULT 47
W16759
LOCUS   zbl6f04.r1 Soares_fetal_lung_NbHL19w Homo sapiens cDNA clone
DEFINITION IMAGE:302239 5' similar to contains element MSR1 repetitive element
; mRNA sequence.
ACCESSION W16759.1 GI:1291140
VERSION   W16759
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

REFERENCE 1 (bases 1 to 517)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
Wilson,R.
TITLE     The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: FPrimer
High quality sequence stop: 348.
FEATURES
    source
        1..517
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="GDB:1247171"
            /db_xref="taxon:9606"
            /clone="IMAGE:302239"
            /dev_stage="19 weeks"
            /lab_host="DH10B (ampicillin resistant)"
            /clone_lib="Soares_fetal_lung_NbHL19w"
            /note="Organ: lung; Vector: pT7T3D (Pharmacia) with a
            modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
            strand cDNA was primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAATGGAGCGCGCGCAATTTTTTTTTTTTTTTT-3'],
            double-stranded cDNA was size selected, ligated to Eco RI
            adapters (Pharmacia), digested with Not I and cloned into
            the Not I and Eco RI sites of a modified pT7T3 vector
            (Pharmacia). Library went through one round of
            normalization to a Cot = 5. Library constructed by Bento
            Soares and M.Fatima Bonaldo. This library was constructed
            from the same focus as the fetal heart library, Soares
            fetal heart NBHL19w."

ORIGIN
Query Match 11.4%; Score 274; DB 14; Length 517;
Best Local Similarity 99.7%; Pred. No. 1.2e-128;
Matches 394; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY      1159 ACTGGTTTAAATGCCAGTCGGCCATTCTCTGTACATAAAAGCTAGCAATGGCGCTC 1218
Db      1 ACTGGTTTAAATGCCAGTCGGCCATTCTCTGTACATAAAAGCTAGCAATGGCGCTC 60
QY      1219 TGGTATCAATGAATGACTCATTGTATCCAGTCATATTAGATCGGTACTCAGCCAA 1278

```

```

Db      61 TGGTATCAATGAATGACTCATTGTATCTACCAAGTGATATTAGATCGGTACTCAGCCAA 120
QY      1279 CAAGCCTATGTGCTCTTTTATATCAGGTCCCATCATGTGAAAAATGGAGGTGAACTTACT 1338
Db      121 CAAGCCTATGTGCTCTTTTATATCAGGTCCCATCATGTGAAAAATGGAGGTGAACTTACT 180
QY      1339 CATCCACCCATAGCCCGCGCCAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGGGTTGTC 1398
Db      181 CATCCACCCATAGCCCGCGCCAGTCTCTCTCCCGCCCGCTCATCAGTCAGCGGGTTGTC 240
QY      1399 ACCAAACAAACAGGCTGCCGAGGCTTTATCGGACCACAGCTTCCTCTCATATGATAAAG 1458
Db      241 ACCAAACAAACAGGCTGCG-CAGGCTTTATCGGACCACAGCTTCCTCTCATATGATAAAG 299
QY      1459 AATCCACCTCAGTTAAATGGGACTGGACCATTTGAAGACACGCGACGCTTCATGTCG 1518
Db      300 AATCCACCTCAGTTAAATGGGACTGGACCATTTGAAGACACGCGACGCTTCATGTCG 359
QY      1519 AGTCCTAAACGGGAATTCAGTGTCAACAGGGCTAG 1553
Db      360 AGTCCTAAACGGGAATTCAGTGTCAACAGGGCTAG 394

RESULT 48
AA749124/c
LOCUS   AA749124
DEFINITION my03e05.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1270688 3',
mRNA sequence.
ACCESSION AA749124
VERSION   AA749124.1 GI:27899082
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens

REFERENCE 1 (bases 1 to 389)
AUTHORS   Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 875 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 370.
FEATURES
    source
        1..389
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="IMAGE:1270688"
            /tissue_type="germinal center B cell"
            /lab_host="DH10B"
            /clone_lib="NCI_CGAP_GCB1"
            /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
            polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
            was prepared from human tonsillar cells enriched for
            germinal center B cells by flow sorting (CD20+, IgD-,
            provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
            (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
            primed with a Not I - oligo(dT) primer
            [5'-TGTTACCAATCTGAATGGAGCGCGCTCATTTTTTTTTTTTTTTT-3',
            ]. Double-stranded cDNA was ligated to Eco RI adaptors
            (Pharmacia), digested with Not I and cloned into the Not I

```

and Eco RI sites of the modified p77T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN
Query Match 11.1%; Score 267; DB 9; Length 389;
Best Local Similarity 99.7%; Pred. No. 4.4e-125;
Matches 387; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1943 AGGAGATGGGATGGTACGATTGTGAGCTCCACATCTCCCGCCCAAGATGCCGAGATG 2002
Db 389 AGGAGATGGGATGGTACGATTGTGAGCTCCACATCTCCCGCCCAAGATGCCGAGATG 330
QY 2003 AGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAAACGGTGTATAGTGCAG 2062
Db 329 AGGAGGCCACTCCGACGAGCTTCAAGAACCCATGACCCCTAAACGGTGTATAGTGCAG 270
QY 2063 ACAGGACAGTACCCGAAAGAAACGGCTAGCGCTGATGTG-CCAGCTGCCAAGGC 2121
Db 269 ACAGGACAGTACCCGAAAGAAACGGCTAGCGCTGATGTGCGCCAGCTGCCAAGGC 210
QY 2122 CAGCGTCCCTGACCTCAGAAAATCCCTTTGTCTAAGGCAAAACGGTCTTCTCGAAAGTTG 2181
Db 209 CAGCGTCCCTGACCTCAGAAAATCCCTTTGTCTAAGGCAAAACGGTCTTCTCGAAAGTTG 150
QY 2182 ATGCGTCTCTCTTGTCTCTCCAGAAAGCAAAATCTTAGAGACCTTCAGGCTTAGC 2241
Db 149 ATGCGTCTCTCTTGTCTCTCTCCAGAAAGCAAAATCTTAGAGACCTTCAGGCTTAGC 90
QY 2242 AACAACTGAAGGCTGACGGATGAATGATGTCACCTGAGCAGAGAGGGGCCCTCCC 2301
Db 89 AACAACTGAAGGCTGACGGATGAATGATGTCACCTGAGCAGAGAGGGGCCCTCCC 30
QY 2302 GAGGACCGGACGCGCGGCTCAGCCTG 2329
Db 29 GAGGACCGGACGCGCGGCTCAGCCTG 2

RESULT 49
W19122 476 bp mRNA linear EST 03-MAY-1996
LOCUS zbl5b04.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:302095 5' similar to contains element MSR1 repetitive element
i, mRNA sequence.
W19122
W19122.1 GI:1294865
EST.
Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ETPRimer
High quality sequence stop: 373.
FEATURES
1. .476
/organism="Homo sapiens"
/mol_type="mRNA"

ORIGIN
Query Match 10.9%; Score 264; DB 14; Length 476;
Best Local Similarity 99.7%; Pred. No. 1.6e-123;
Matches 384; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1159 ACTGGTTTAAATGGCCATGCTGGCCATTAATCTCTGTACATAAAAGCTAGCAATGGCCTC 1218
Db 1 ACTGGTTTAAATGGCCATGCTGGCCATTAATCTCTGTACATAAAAGCTAGCAATGGCCTC 60
QY 1219 TGTATCAATGAATGACTCCATTTATATCAGGTCCCATGATGTGAAAATGAGGTGAATCTTACT 1278
Db 61 TGTATCAATGAATGACTCCATTTATATCAGGTCCCATGATGTGAAAATGAGGTGAATCTTACT 180
QY 1279 CAAGCTTATCTGCTCTTTTATATCAGGTCCCATGATGTGAAAATGAGGTGAATCTTACT 1338
Db 121 CAAGCTTATGCTCTTTTATATCAGGTCCCATGATGTGAAAATGAGGTGAATCTTACT 180
QY 1339 CATCCCAACCATAGCCCCGCCAGTCTCTCCCGCCCGCTCATCATCAGCGGGTGTGTC 1398
Db 181 CATCCCAACCATAGCCCCGCCAGTCTCTCCCGCCCGCTCATCATCAGCGGGTGTGTC 240
QY 1399 ACACAAACAGCTGCGCCAGGCTTTATCGGACACAGCTTCCCTCTCACATGATAAG 1458
Db 241 ACCAAACAAACAGCTGCGG-CAGGCTTTATCGGACACAGCTTCCCTCTCACATGATAAG 299
QY 1459 AATCCACTCTCACTTAAATGGGACTGGACCATTTGAAGACACGCCCAAGCAGTTCATGTCG 1518
Db 300 AATCCACTCTCACTTAAATGGGACTGGACCATTTGAAGACACGCCCAAGCAGTTCATGTCG 359
QY 1519 AGTCCTAAACGGAA"TTCCAGTGTCA 1543
Db 360 AGTCCTAAACGGAA"TTCCAGTGTCA 384
RESULT 50
W16787 390 bp mRNA linear EST 29-APR-1996
LOCUS zbl6g05.r1 Soares fetal lung NbHL19W Homo sapiens cDNA clone
DEFINITION IMAGE:302264 5' similar to contains element MSR1 repetitive element
i, mRNA sequence.
W16787
W16787.1 GI:1291168
EST.
Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 390)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ETPRimer
High quality sequence stop: 373.
FEATURES
1. .476
/organism="Homo sapiens"
/mol_type="mRNA"

JOURNAL
COMMENT

Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ETPRimer
High quality sequence stop: 312.

FEATURES

source
1..390
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:1247196"
/db_xref="taxon:9606"
/clone="IMAGE:302284"
/dev_stage="19 weeks"
/lab_host="PH10B (ampicillin resistant)"
/clone_lib="Soares fetal lung NbHL19W"
/note="Organ: lung; Vector: pT7f3D (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer
15'-TGTTACCAATCTGAAGTCGAGCGCGCAATTTTCTTTT-3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7f3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. This library was constructed
from the same fetus as the fetal heart library, Soares
fetal heart NbHL19W."

ORIGIN

Query Match 10.7%; Score 259; DB 14; Length 390;
Best Local Similarity 100.0%; Pred.No. 5.6e-121;
Matches 259; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1159 ACTGGTTTAAATGGCATGCGCCATTACTTCTGCTACATAAAGTAGCAATGGCCTC 1218
Db 1 ACTGGTTTAAATGGCATGCGCCATTACTTCTGCTACATAAAGTAGCAATGGCCTC 60
Qy 1219 TGGTATCAAAATGAATGACTCCATGTCATGATATGATCGTACTCAGCCAA 1278
Db 61 TGGTATCAAAATGAATGACTCCATGTCATGATATGATCGTACTCAGCCAA 120
Qy 1279 CAAGCCTATGTCCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGGAGGTGAATTTACT 1338
Db 121 CAAGCCTATGTCCTCTTTTATATCAGGTCCTCATGATGTGAAAAATGGAGGTGAATTTACT 180
Qy 1339 CATCCACCCATAGCCCGGCGAGTCCTCTCCCGCCCGTCATCAGTCAGGGGTGTC 1398
Db 181 CATCCACCCATAGCCCGGCGAGTCCTCTCCCGCCCGTCATCAGTCAGGGGTGTC 240
Qy 1399 ACCAACAAACAGGTGCGC 1417
Db 241 ACCAACAAACAGGTGCGC 259

Search completed: August 17, 2004, 07:20:10
Job time : 9159 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 16, 2004, 23:48:34 ; Search time 198 Seconds
(without alignments)
6760.311 Million cells used

Title: US-10-049-745-31

reflect score: 2412
 Sequence: 1 gggctgtgcggcgggc.....agggttaattttagcttga 2412

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 30

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_NA:*

```
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
```

```
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
```

5: /cgn2_6/ptodata/2/ina/pctus COMB.se

```
0: /cgn2_b/ptodata/2/1na/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query No.	%			DB	ID	Description
		Score	Match	Length			
C	1	357	14.8	722	4	US-09-280-116-223	Sequence 223, App
	2	205	8.5	457	4	US-09-280-116-214	Sequence 214, App
	3	30	1.2	847	4	US-09-495-050A-40	Sequence 40, Appl
	4	30	1.2	1273	4	US-09-976-594-448	Sequence 448, App

ALIGNMENTS

RESULT 1

```

US-09-280-116-223
; Sequence 223, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/178965
; CURRENT APPLICATION NUMBER: US/09/280.116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 223
; LENGTH: 722
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2

```

```

;
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(722)
; OTHER INFORMATION: n =
US-09-280-116-223

```

Query Match 14.8%; Score 357; DB 4; Length 722;
Best Local Similarity 100.0%; Pred. No. 9.7e-173;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1061	AGGATGTGAATAACCTTGAGTATCTTGATATTCGGCCATATATGTCTCAACCCAAACGGAG	1120
Db	107	AGGATGTGAATAACCTTGAGTATCTTGATATTCGGCCATATATGTCTCAACCCAAACGGAG	166
QY	1121	AGCCAATTGCTACGTCCTTGATGCGAGTGCCTGCTCCACACTGGTTTTAAATGCCATGCTG	1180
Db	167	AGCCAATTGCTACGTCCTTGATGCGAGTGCCTGCTCCACACTGGTTTTAAATGCCATGCTG	226
QY	1181	GCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTCAAAATCAATGACTCCCA	1240
Db	227	GCCATTACTTCTGCTACATAAAAGCTAGCAATGGCCCTCTGGTCAAAATCAATGACTCCCA	286
QY	1241	TTGTATCTACAGTGAATTAGATCGGTACTCAGCCAAAGAGCTATGTCTCTTTTATA	1300
Db	287	TTGTATCTACAGTGAATTAGATCGGTACTCAGCCAAAGAGCTATGTCTCTTTTATA	346
QY	1301	TCAGGTCCCATGATGTGAAAAATGGAGGTGAACCTTACTCCACCCCATAGCCCCGGCC	1360
Db	347	TCAGGTCCCATGATGTGAAAAATGGAGGTGAACCTTACTCCACCCCATAGCCCCGGCC	406
QY	1361	AGTCCTCTCCCCGCCCGTCATCAGTCAGCGGGTTGTACCAACAAACAGGCTGGCG	1417
Db	407	AGTCCTCTCCCCGCCCGTCATCAGTCAGCGGGTTGTACCAACAAACAGGCTGGCG	463

RESULT 2

```

US-09-280-116-214
; Sequence 214, Application US/09280116A
; Patent No. 6331427
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs
; FILE REFERENCE: 5800-24, 035800/176965
; CURRENT APPLICATION NUMBER: US/09/280,116A
; CURRENT FILING DATE: 1999-03-26
; NUMBER OF SEQ ID NOS: 268
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 214
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: ubiquitin carboxyl-terminal hydrolases family 2
US-09-280-116-214

```

Query Match	8.5%;	Score 205;	DB 4;	Length 457;
Best Local Similarity	100.0%;	Pred. No. 7.3e-95;		
Matches 205;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	138	CTCCGAGGCAGTCTC	CTCGAGACATGG	ATGAGTTC	TGCCAGCTGGGGTGC	TGTGTGTC	197
Db	103	CTCCGAGGCAGTCTC	CTCGAGACATGG	ATGAGTTC	TGCCAGCTGGGGTGC	TGTGTGTC	162
QY	198	TTCAATTGAATGATGTG	TCAAATCACACACT	TTCTTTTAGGAC	CAGTACCTGGTGTGTG	TGATG	257
Db	163	TTCAATTGAATGATGTG	TCAAATCACACACT	TTCTTTTAGGAC	CAGTACCTGGTGTGTG	TGATG	222
QY	258	TTATTTCGAGTTTCAT	CTGTACTGATAAAT	CAAAACCATCACC	ACAAAAGGATCAAGCCCT	317	
Db	223	TTATTTCGAGTTTCAT	CTGTACTGATAAAT	CAAAACCATCACC	ACAAAAGGATCAAGCCCT	282	
QY	318	AGGTGATGGCATCGCT	CCCTCCACAG	342			

Wed Aug 18 13:52:59 2004

Db 283 AGGTGATGCGATCGCTCTCCACAG 307

RESULT 3

US-09-495-050A-40/c
; Sequence 40, Application US/09495050A
; Patent No. 6492505
; GENERAL INFORMATION:
; APPLICANT: Roopa, Reddy
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P
; FILE REFERENCE: PA-0013 US
; CURRENT APPLICATION NUMBER: US/09/495,050A
; CURRENT FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/118,318
; PRIOR FILING DATE: February 1, 1999
; NUMBER OF SEQ ID NOS: 305
; SOFTWARE: PERL Program
; SEQ ID NO 40
; LENGTH: 847
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6492505 797777CTT
US-09-495-050A-40

Query Match 1.2%; Score 30; DB 4; Length 847;
Best Local Similarity 100.0%; Pred. No. 3.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 2412
Db 37 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 8

RESULT 4

US-09-976-594-448
; Sequence 448, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 448
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 522433CB1
US-09-976-594-448

Query Match 1.2%; Score 30; DB 4; Length 1273;
Best Local Similarity 100.0%; Pred. No. 3.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 2412
Db 1238 CCCTTTAGTCAGGGTTAAATTTTAGCTTGCA 1267

Search completed: August 17, 2004, 07:23:31
Job time : 201 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 17, 2004, 04:45:41 ; Search time 3503 Seconds
(without alignments)
3378.458 Million cell updates/sec

Title: US-10-049-745-31

Perfect score: 2412

Sequence: 1 gggctgtgcggcgccg.....agggttaatttagcttgca 2412

Scoring table:

Gapop_60.0 , Gapext 60.0

Searched: 3225727 seqs, 2453303834 residues

Word size : 30

Total number of hits satisfying chosen parameters: 34

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/2/pubpna/US05_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2321	96.2	4659	13	US-10-399-645-22
2	2287	94.8	2347	15	US-10-163-547-2
3	2249	93.2	3951	16	US-10-379-981-8
4	2210	91.6	4483	13	US-10-302-172-390
5	333	13.8	333	9	US-09-864-761-31553
6	314	13.0	498	9	US-09-864-761-15021
7	205	8.5	489	9	US-09-864-761-14092
8	198	8.2	198	9	US-09-864-761-30650
9	157	6.5	504	13	US-10-027-632-128282
10	157	6.5	504	16	US-10-027-632-128282
11	97	4.0	631	13	US-10-027-632-277717
12	97	4.0	631	16	US-10-027-632-277717
13	68	2.8	591	13	US-10-027-632-85072
14	68	2.8	591	16	US-10-027-632-85072

15	41	1.7	3981	16	US-10-379-981-11	Sequence 11, Appl
16	39	1.6	592	13	US-10-027-632-319601	Sequence 319601,
17	39	1.6	592	16	US-10-027-632-319601	Sequence 319601,
18	30	1.2	637	17	US-10-258-662-51	Sequence 51, Appl
19	30	1.2	733	15	US-10-149-819-39	Sequence 39, Appl
20	30	1.2	847	15	US-10-313-542-40	Sequence 40, Appl
21	30	1.2	972	9	US-09-349-015-17	Sequence 17, Appl
22	30	1.2	972	15	US-10-219-664-13	Sequence 13, Appl
23	30	1.2	1112	9	US-09-880-192-16	Sequence 16, Appl
24	30	1.2	1112	15	US-10-427-348-16	Sequence 16, Appl
25	30	1.2	1273	9	US-09-981-353-187	Sequence 187, App
26	30	1.2	1273	10	US-09-919-039-191	Sequence 191, App
27	30	1.2	1273	15	US-10-133-013-59	Sequence 59, Appl
28	30	1.2	1702	9	US-09-349-015-2	Sequence 2, Appl
29	30	1.2	1702	15	US-10-219-664-2	Sequence 2, Appl
30	30	1.2	1833	9	US-09-880-192-32	Sequence 32, Appl
31	30	1.2	1833	15	US-10-427-348-32	Sequence 32, Appl
32	30	1.2	1826	10	US-09-971-392-216	Sequence 216, App
33	30	1.2	1987	9	US-09-818-143-20	Sequence 20, Appl
34	30	1.2	2040	15	US-10-198-846-10830	Sequence 10830, A

ALIGNMENTS

RESULT 1

US-10-399-645-22
; Sequence 22, Application US/10399645
; Publication No. US20040029249A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE CORPORATION; LEE, Ernestine A.
; APPLICANT: HAFALIA, April J.A.; YUE, Henry
; APPLICANT: LAL, Preeti G.; YAO, Monique G.
; APPLICANT: LU, Yan; CHAWLA, Narinder K.
; APPLICANT: WARREN, Bridget A.; LU, Dyung Aina M.
; APPLICANT: BAUGHN, Mariah R.; DELEGANE, Angelo M.
; APPLICANT: BURFORD, Neil; BOROWSKY, Mark L.
; APPLICANT: LEE, Sally; XU, Yuming
; APPLICANT: GRIFFIN, Jennifer A.; KALLICK, Deborah A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: ISON, Craig H.; TANG, Y. Tom
; APPLICANT: AZIMZAI, Valda; ELLIOTT, Vicki S.
; APPLICANT: SWARNAKAR, Anita; RAMKUMAR, Jayalaxmi
; APPLICANT: NGUYEN, Daniel B.; TRIBOULEY, Catherine M.
; APPLICANT: LO, Terence P.; AU-YOUNG, Janice K.
; APPLICANT: THANGAVELU, Kavitha; KEARNEY, Liam
; TITLE OF INVENTION: PROTEASES
; FILE REFERENCE: PI-0263 USN
; CURRENT APPLICATION NUMBER: US/10/399,645
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: PCT/US01/51034
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: US 60/241,573
; PRIOR FILING DATE: 2000-10-18
; PRIOR APPLICATION NUMBER: US 60/243,643
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/245,256
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/248,395
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US 60/249,826
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: US 60/252,303
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/250,981
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 4659
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:

i	NAME/KEY: misc_feature
i	OTHER INFORMATION: Incyte ID No. US20040029249A1 7997881CB1
US-10-399-645-22	
	Query Match 96.2%; Score 2321; DB 13; Length 4659;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2371; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY	1 GGGCTGTGTCGGCGCGCGCGCGCGCGAGAGGGGATGGAGCGAGCGCCGACCGGGT 60
Db	83 GGGCTGTGTCGGCGCGCGCGCGCGAGAGGGGATGGAGCGAGCGCCGACCGGGT 142
QY	61 CAGAGTTGAACAATGACCACTAGTGACAAAGCTTCTGAATCTTCAGAGCCCATCAGCCTAT 120
Db	143 CAGAGTTGAACAATGACCATTAGTGACAAAGCTTCTGAATCTTCAGAGCCCATCAGCCTAT 202
QY	121 CAGAATCAGCTGGCAGCTCCGAGGCAGTCTCACCTGGAGACATGGATGCAAGTTCGCC 180
Db	203 CAGAATCAGCTGGCAGCTCCGAGGCAGTCTCACCTGGAGACATGGATGCAAGTTCGCC 262
QY	181 AGCTGGGGTCTGTGCTTCATTTGAANGATGTGTAATTCACACTTTCCTTAGGACCA 240
Db	263 AGCTGGGGTCTGTGCTTCATTTGAATGATGTCAAATTCACACTTTCCTTAGGACCA 322
QY	241 GTACCTGGTCTGTAGTTTTTTCAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA 300
Db	323 GTACCTGGTCTGTAGTTTTTTCAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA 382
QY	301 CAAAAGGATCAAGCCCTTAGTGATGTCATCGCTCTCCACAGAAAGTTCCTTTCCCATCT 360
Db	383 CAAAAGGATCAAGCCCTTAGTGATGTCATCGCTCTCCACAGAAAGTTCCTTTCCCATCT 442
QY	361 GAGAAGATTTGTCTTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 420
Db	443 GAGAAGATTTGTCTTTAAGTGGCAACAACTCATAGAGTTGGAGCTGGGCTCCAGAAATTG 502
QY	421 GGCAATACCTGTTTGGCCAATGACGACTGCAGTGTTTAACCTTACACACCACTCTTGCC 480
Db	503 GGCAATACCTGTTTGGCCAATGACGACTGCAGTGTTTAACCTTACACACCACTCTTGCC 562
QY	481 AATTACATGCTATCACATGAACACTCCAAACATGTCATGAGAGGCTTTTGTATGATG 540
Db	563 AATTACATGCTATCACATGAACACTCCAAACATGTCATGAGAGGCTTTTGTATGATG 622
QY	541 TGTACAAATGCAAGCACATATPACCCAGGCACTCAGTAATCCTGGGAGCGTTATTAAACCA 600
Db	623 TGTACAAATGCAAGCACATATPACCCAGGCACTCAGTAATCCTGGGAGCGTTATTAAACCA 682
QY	601 ATGTTTGTTCATCAATGAGATGCGCGGTATAGCTAGGCACTCCGTTTGGAAACCAAGAA 660
Db	683 ATGTTTGTTCATCAATGAGATGCGCGGTATAGCTAGGCACTCCGTTTGGAAACCAAGAA 742
QY	661 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGCC 720
Db	743 GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAGAAAGCATGCTTGAATGCC 802
QY	721 AGCAATAAATTTAGACAGACACACCCAGSCCACTCTTGTGTTGCAGATATTGAGGGA 780
Db	803 AGCAATAAATTTAGACAGACACACCCAGSCCACTCTTGTGTTGCAGATATTGAGGGA 862
QY	781 TACCTAAGANCTAGAGTCAAAATGTTTAAATTTGCAAGGGCGTTTCAGATATCTTTGATCCA 840
Db	863 TACCTAAGATCTAGAGTCAAAATGTTTAAATTTGCAAGGGCGTTTCAGATATCTTTGATCCA 922
QY	841 TATCTTTGATATAACATTCGAGATAAAGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAG 900
Db	923 TATCTTTGATATAACATTCGAGATAAAGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAG 982
QY	901 TTTTGTCAAGCCGGAACAGCTTTCATGCGAGAAAACTCGTACAAAGTCGACGAGTGTAAAG 960
Db	983 TTTTGTCAAGCCGGAACAGCTTTCATGCGAGAAAACTCGTACAAAGTCGACGAGTGTAAAG 1042
QY	961 AUGGTTCCAGCTTCAAGAGGTTCACTATCCATAGATCCTCTAATGTCTTTACACTTTCT 1020

QY 493 TCACATGAACACCTCCAAAAACATGTCATGCAAGGCTTTTGTATGATGTGTACAAATGCAA 552
Db 421 TCACATGAACACCTCCAAAAACATGTCATGCAAGGCTTTTGTATGATGTGTACAAATGCAA 480
QY 553 GCACATATTACCCAGGACACTCAGTAATCTCTGGGACAGTATTAAACCAATGTTGTGTCATC 612
Db 481 GCACATATTACCCAGGACACTCAGTAATCTCTGGGACAGTATTAAACCAATGTTGTGTCATC 540
QY 613 AATGAGATGGGGGTATAGCTAGGCACCTCGTTTGGAAACCAAGAGATGCCATGAA 672
Db 541 AATGAGATGGGGGTATAGCTAGGCACCTCGTTTGGAAACCAAGAGATGCCATGAA 600
QY 673 TTCCTTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAGCAATAAATTA 732
Db 601 TTCCTTCAATACACTGTTGATGCTATGACAGAAAGCATGCTTGAATGGCAGCAATAAATTA 660
QY 733 GACAGACACACCCAGGACCACTCTGTTGTCAGATATTGGAGGATACCTTAAGATCT 792
Db 661 GACAGACACACCCAGGACCACTCTGTTGTCAGATATTGGAGGATACCTTAAGATCT 720
QY 793 AGAGTCAAAATGTTTAAATGCAAGGCGTTCAGATACCTTTGATCCATCTCTGATATA 852
Db 721 AGAGTCAAAATGTTTAAATGCAAGGCGTTCAGATACCTTTGATCCATCTCTGATATA 780
QY 853 ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTCGTGAAGCCG 912
Db 781 ACATTGGAGATAAAGGCTGCTCAGAGTGTCAACAAGGCATTGGAGCAGTTCGTGAAGCCG 840
QY 913 GAAACAGCTTGATGGAGAAACTCGTACAGTGCAGCAAGTGTAAAGATGTTTCCAGCT 972
Db 841 GAAACAGCTTGATGGAGAAACTCGTACAGTGCAGCAAGTGTAAAGATGTTTCCAGCT 900
QY 973 TCAAGAGAGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTCTCTGAAACGTTTTT 1032
Db 901 TCAAGAGAGTTCACTATCCATAGATCCTCTAATGTTCTTACACTTCTCTGAAACGTTTTT 960
QY 1033 GCAATTTTACCGGTGGAAAAATTCCTAAGGATGTGAAATACCTCAGTATCTTGATATT 1092
Db 961 GCAATTTTACCGGTGGAAAAATTCCTAAGGATGTGAAATACCTCAGTATCTTGATATT 1020
QY 1093 CGGCATATATGTCACACCAACGAGAGCCAAATGTCAGTCTTGATGCAAGTCTG 1152
Db 1021 CGGCATATATGTCACACCAACGAGAGCCAAATGTCAGTCTTGATGCAAGTCTG 1080
QY 1153 GTCCACACTGGTTTAAATGGCATCTGCGCAATTAATCTGCTACATAAAAGCTAGCAAT 1212
Db 1081 GTCCACACTGGTTTAAATGGCATCTGCGCAATTAATCTGCTACATAAAAGCTAGCAAT 1140
QY 1213 GGCCTCTGGTATCAAAATGAATGATCCATTTGATCTACAGTGTATATAGATCGGTACTC 1272
Db 1141 GGCCTCTGGTATCAAAATGAATGATCCATTTGATCTACAGTGTATATAGATCGGTACTC 1200
QY 1273 AGCCAAACAGCCTATGTCCTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAA 1332
Db 1201 AGCCAAACAGCCTATGTCCTTTATATCAGTCCCATGATGTGAAAAATGGAGGTGAA 1260
QY 1333 CTCTACTATCCCAACCCATAGCCCGGCGCAGTCTCTCCCGCCCGCATCAGTCAGCGG 1392
Db 1261 CTCTACTATCCCAACCCATAGCCCGGCGCAGTCTCTCCCGCCCGCATCAGTCAGCGG 1320
QY 1393 GTTGTACCAACAAACAGGCTGCGCCAGGCTTTATCGSACCAACAGCTTCCCTCTCACATG 1452
Db 1321 GTTGTACCAACAAACAGGCTGCGCCAGGCTTTATCGSACCAACAGCTTCCCTCTCACATG 1380
QY 1453 ATAAGAATCCACCTCACTTAATGGGACTGGACCAATGAAAGACAGCCCAAGCAGTTCC 1512
Db 1381 ATAAGAATCCACCTCACTTAATGGGACTGGACCAATGAAAGACAGCCCAAGCAGTTCC 1440
QY 1513 ATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGCT 1572
Db 1441 ATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGCTAGTCTGTTAATGCTTCAGCT 1500
QY 1573 TCTGTCCAAAACTGGTCAGTTAATAGTCTCCTCAGTATCCAGAACATCCTTAAGAACAA 1632

Db 1501 TCTGTCCAAAACTGGTCAGTTAATAGTCTCTCAGTATCCAGAACATCTCTTAAGAACAA 1560
QY 1633 AAAATTACAATCAGTATTCAACAAGTTCGCTGTTGCGCAGTGTCACTCTCAACCTTAAC 1692
Db 1561 AAAATTACAATCAGTATTCAACAAGTTCGCTGTTGCGCAGTGTCACTCTCAACCTTAAC 1620
QY 1693 CTTCAATAGTAAATCTTTGGAGAACCTTACCAAGCCCTTCCCTCTTCTTACCATTAACCAAT 1752
Db 1621 CTTCAATAGTAAATCTTTGGAGAACCTTACCAAGCCCTTCCCTCTTCTTACCATTAACCAAT 1680
QY 1753 TCTGCACTACAGTCTTACCTGAAAGCATCTAOCATGTCAGTTCCTAGTAAAGTAAACAAA 1812
Db 1681 TCTGCACTACAGTCTTACCTGAAAGCATCTAOCATGTCAGTTCCTAGTAAAGTAAACAAA 1740
QY 1813 CCGATCCCGCGCAGTGAATCTCTCCAGCCGCTGATGATGATGATGATGATGATGATGATGAT 1872
Db 1741 CCGATCCCGCGCAGTGAATCTCTCCAGCCGCTGATGATGATGATGATGATGATGATGATGAT 1800
QY 1873 TCAGAGTCTGCTGCTGCTTATGGCGCGGCTCTCTGAGGACTCTGACGAGGAGTCAAG 1932
Db 1801 TCAGAGTCTGCTGCTGCTTATGGCGCGGCTCTCTGAGGACTCTGACGAGGAGTCAAG 1860
QY 1933 GGGCTGGGCAAGGAGAAATGGGATTTGGTACGATTTGAGCTTCCACTCTCCCGGCCAAGAT 1992
Db 1861 GGGCTGGGCAAGGAGAAATGGGATTTGGTACGATTTGAGCTTCCACTCTCCCGGCCAAGAT 1920
QY 1993 GCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGATGATGATGATGATGATGAT 2052
Db 1921 GCCGAAGATGAGGAGGCCACTCCGACGAGCTTCAAGAACCCATGATGATGATGATGATGATGAT 1980
QY 2053 AATAGTGCAGACAGCAGACAGTGCACGAGCTTCAAGAACCCATGATGATGATGATGATGATGAT 2112
Db 1981 AATAGTGCAGACAGCAGACAGTGCACGAGCTTCAAGAACCCATGATGATGATGATGATGATGAT 2040
QY 2113 TSCCAAGGCCAGCTTCCCTGCACTCAGAAAATCCCTTTGCTAAGGCAACCGTCTTCTCT 2172
Db 2041 TSCCAAGGCCAGCTTCCCTGCACTCAGAAAATCCCTTTGCTAAGGCAACCGTCTTCTCT 2100
QY 2173 GGAAGTTGATGCTGCTTCTTCTGCTGCTTCTGCTGCTTCTGCTGCTTCTGCTGCTTCTGCT 2232
Db 2101 GGAAGTTGATGCTGCTTCTTCTGCTGCTTCTGCTGCTTCTGCTGCTTCTGCTGCTTCTGCT 2160
QY 2233 AGGCTTAGCAACAACTGAAAGGCTCGACGGATGAAATGAGTGACCTGGAGCAGAGG 2292
Db 2161 AGGCTTAGCAACAACTGAAAGGCTCGACGGATGAAATGAGTGACCTGGAGCAGAGG 2220
QY 2293 GGCCTTCCCGAGGACCGCAGCGGAGCCTCAGCTGGCAGCCCGCCGCAATCCCTG 2352
Db 2221 GGCCTTCCCGAGGACCGCAGCGGAGCCTCAGCTGGCAGCCCGCCGCAATCCCTG 2280
QY 2353 GAGGAGCCAGATGCGGCCG 2372
Db 2281 GAGGAGCCAGATGCGGCCG 2300

RESULT 4

US-10-302-172-390
; Sequence 390, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Drmanac, Radong J.
; TITLE OF INVENTION: No. US20040053250A1 Arginine-rich Protein-like Nucleic Acids and
; FILE REFERENCE: 803_ICNCP
; CURRENT APPLICATION NUMBER: US/10/302,172
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/225,251
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: PCT US02/05095
; PRIOR FILING DATE: 2002-03-05

; PRIOR APPLICATION NUMBER: US 09/799,451
 ; PRIOR FILING DATE: 2001-03-05
 ; NUMBER OF SEQ ID NOS: 950
 ; SOFTWARE: pt_FL_genes Version 2.0
 ; SEQ ID NO 390
 ; LENGTH: 4483
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (191)..(4000)
 ; US-10-302-172-390

Query Match 91.6%; Score 2210; DB 13; Length 4483;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2310; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	61	CAGAGTTGAACATGACCATAGTTGACAAAGCTTCTGAAATCTTCAGACCCATCAGCCTAT	120
Db	179	CAGAGTTGAACATGACCATAGTTGACAAAGCTTCTGAAATCTTCAGACCCATCAGCCTAT	238
QY	121	CAGAAATCAGCTCGCAGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGCTTCTGCC	180
Db	239	CAGAAATCAGCTCGCAGCTCCGAGGAGTCTCACCTGGAGACATGGATGAGGCTTCTGCC	298
QY	181	AGCTGGGGTCTGTGCTTCAATGAATGATGTGTCAAAATCACACATTTCTTTAGGACCA	240
Db	299	AGCTGGGGTCTGTGCTTCAATGAATGATGTGTCAAAATCACACATTTCTTTAGGACCA	358
QY	241	GTACCTGGTCTGTAGTTTATTCAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA	300
Db	359	GTACCTGGTCTGTAGTTTATTCAGTTTCATCTGTACCTGATAAATCAAAACCATCACCA	418
QY	301	CAAAAGGATCAAGCCCTAGGTGATGATCGCTCCTCCACAGAAAGTTCTTTTCCCATCT	360
Db	419	CAAAAGGATCAAGCCCTAGGTGATGATCGCTCCTCCACAGAAAGTTCTTTTCCCATCT	478
QY	361	GAGAAGATTTGCTTAAAGTGACCAAAATCATAGATGTGGAGTGGGCTCCAGAAATTG	420
Db	479	GAGAAGATTTGCTTAAAGTGACCAAAATCATAGATGTGGAGTGGGCTCCAGAAATTG	538
QY	421	GGCAATACCTGTTTGGCAATGACGACCTGAGGTTTAACTACACACACCTCTTGCC	480
Db	539	GGCAATACCTGTTTGGCAATGACGACCTGAGGTTTAACTACACACACCTCTTGCC	598
QY	481	AATTACATGCTATCACATGAACACTCCAAACATGTCTATGAGAGGCTTTTGTATGATG	540
Db	599	AATTACATGCTATCACATGAACACTCCAAACATGTCTATGAGAGGCTTTTGTATGATG	658
QY	541	TGTAAATGCAAGCACATATTACCCAGGACCTCAGTAATCTCTGGGACGTTTAAACCA	600
Db	659	TGTAAATGCAAGCACATATTACCCAGGACCTCAGTAATCTCTGGGACGTTTAAACCA	718
QY	601	ATGTTTGTATCAATGAGATGGGGTATAGTGGGACCTCCGTTTGGAAACCAAGAA	660
Db	719	ATGTTTGTATCAATGAGATGGGGTATAGTGGGACCTCCGTTTGGAAACCAAGAA	778
QY	661	GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAAGAAAGCATGTTGAATGCG	720
Db	779	GATGCCCATGAATTCCTTCAATACACTGTTGATGCTATGCAAGAAAGCATGTTGAATGCG	838
QY	721	AGCAATAAATAGACAGACACACCCAGGACCACTCTGTTGTTGTCAGATATTTGGAGGA	780
Db	839	AGCAATAAATAGACAGACACACCCAGGACCACTCTGTTGTTGTCAGATATTTGGAGGA	898
QY	781	TACCTAAGATCTAGAGTCAAAATGTTTAAATTCAGAGGGGTTTTCAGATATTTGATCCA	840
Db	899	TACCTAAGATCTAGAGTCAAAATGTTTAAATTCAGAGGGGTTTTCAGATATTTGATCCA	958
QY	841	TATCTTGTATATAACATTTGGAGATAAAGGCTGTCTCAGAGTGTCAACAGGCATTTGGAGCAG	900
Db	959	TATCTTGTATATAACATTTGGAGATAAAGGCTGTCTCAGAGTGTCAACAGGCATTTGGAGCAG	1018

QY	901	TTTGTGAAGCCGGAACAGCTTGTATGAGAAATCTGCTACAGTGCAGCAAGTGTAAAAAG	960
Db	1019	TTTGTGAAGCCGGAACAGCTTGTATGAGAAATCTGCTACAGTGCAGCAAGTGTAAAAAG	1078
QY	961	ATGGTTCAGCTTCAAGAGGTTCACTATCCATAGATCTCTAAATGTTCTTACACTTTCT	1020
Db	1079	ATGGTTCAGCTTCAAGAGGTTCACTATCCATAGATCTCTAAATGTTCTTACACTTTCT	1138
QY	1021	CTGAAACGTTTTGCAAAATTTTACCGGTGAAAAATTTGCTTAAGGATGTGAAATACCCCTGAG	1080
Db	1139	CTGAAACGTTTTGCAAAATTTTACCGGTGAAAAATTTGCTTAAGGATGTGAAATACCCCTGAG	1198
QY	1081	TATCTTGATATTCGGCCATATATGCTCAACCAACGAGAGCCAAATTCCTACGTTCTG	1140
Db	1199	TATCTTGATATTCGGCCATATATGCTCAACCAACGAGAGCCAAATTCCTACGTTCTG	1258
QY	1141	TATGAGTCTGGTCCACACTGGTTTTTAATGCACTGTGGCCATTTACTTCTGCTACATA	1200
Db	1259	TATGAGTCTGGTCCACACTGGTTTTTAATGCACTGTGGCCATTTACTTCTGCTACATA	1318
QY	1201	AAAGCTAGCAATGGCCTCTGCTGATCAAAATGAATGACTCCATTTGATCTACAGTGATTT	1260
Db	1319	AAAGCTAGCAATGGCCTCTGCTGATCAAAATGAATGACTCCATTTGATCTACAGTGATTT	1378
QY	1261	AGATCGGTACTCAGCAACAAAGCCTATGCTCTTTTATATCAGTCCCATGATGTGAAA	1320
Db	1379	AGATCGGTACTCAGCAACAAAGCCTATGCTCTTTTATATCAGTCCCATGATGTGAAA	1438
QY	1321	AATGGAGTGAACTTACTCATCCACCATAGCCCGGCCAGTCTCTCCCGCCCGCCTC	1380
Db	1439	AATGGAGTGAACTTACTCATCCACCATAGCCCGGCCAGTCTCTCCCGCCCGCCTC	1498
QY	1381	ATCAGTCAGCGGGTGTGTCACCAACAAAGCCTGCGCCAGGCTTTATCGGACCAACAGCTT	1440
Db	1499	ATCAGTCAGCGGGTGTGTCACCAACAAAGCCTGCGCCAGGCTTTATCGGACCAACAGCTT	1558
QY	1441	CCCTCTCACATGATPAAAGATTCACCTCATTAAATGGGACTGACCATTTGAAGACACG	1500
Db	1559	CCCTCTCACATGATPAAAGATTCACCTCATTAAATGGGACTGACCATTTGAAGACACG	1618
QY	1501	CCAAGCAGTTCCATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTT	1560
Db	1619	CCAAGCAGTTCCATGTCAGTCTTAACGGGAATTCAGTGTCAACAGGGCTAGTCTGTT	1678
QY	1561	AATGCTTCAGTCTTCTGTCCAAACTGCTCAGTTAATAGGTCCTCAGTGATCCGAGACAT	1620
Db	1679	AATGCTTCAGTCTTCTGTCCAAACTGCTCAGTTAATAGGTCCTCAGTGATCCGAGACAT	1738
QY	1621	CCTAAGAAACAAAATTTACAAATCAGTATTCAACAAAGTTGCTGTTGCGCAGTGTGAG	1680
Db	1739	CCTAAGAAACAAAATTTACAAATCAGTATTCAACAAAGTTGCTGTTGCGCAGTGTGAG	1798
QY	1681	TCTCAACCTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCT	1740
Db	1799	TCTCAACCTAACCTTCATAGTAATTTCTTTGGAGAACCTTACCAAGCCGTTCCCTCTTCT	1858
QY	1741	ACCATTACCAATTTCTGAGTACAGTCTACCTCGAACGCAATCTACGATGTCTAGT	1800
Db	1859	ACCATTACCAATTTCTGAGTACAGTCTACCTCGAACGCAATCTACGATGTCTAGT	1918
QY	1801	AAAGTAAACAAACCCATCCCGGAGTGAATCTCTCCAGCCGCTGATGAATGGCAAA	1860
Db	1919	AAAGTAAACAAACCCATCCCGGAGTGAATCTCTCCAGCCGCTGATGAATGGCAAA	1978
QY	1861	TCCAGAGTGAATCTCCAGGCTGCTGCTGCTATGGCGCGAGTCTCTGAGGACTCTGAC	1920
Db	1979	TCCAGAGTGAATCTCCAGGCTGCTGCTGCTATGGCGCGAGTCTCTGAGGACTCTGAC	2038
QY	1921	GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATGGTGAATGAGTGTGAGTCTCCACTCT	1980
Db	2039	GAGGAGTCAAAAGGGCTGGGCAAGGAGATGGGATGGTGAATGAGTGTGAGTCTCCACTCT	2098
QY	1981	CCCGGCCAAGATGCGAAGATGAGGAGGCCACTCCGACAGGCTTCAAGAAACCCATGACC	2040


```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15021
; LENGTH: 498
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.66
;
US-09-864-761-15021
;
Query Match 13.0%; Score 314; DB 9; Length 498;
Best Local Similarity 100.0%; Pred. No. 2.1e-158;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1865 AGCTGAACCTCAGCGTGTGTCCTATGGCGCGAGTCTCTGAGGACTCTGACGAGG 1924
Db 498 AGCTGAACCTCAGCGTGTGTCCTATGGCGCGAGTCTCTGAGGACTCTGACGAGG 439
QY 1925 AGTCAAGCGGCTGGCAGGAGATGGATTGTGACGATTTGAGTCCCACTCTCCCG 1984
Db 438 AGTCAAGCGGCTGGCAGGAGATGGATTGTGACGATTTGAGTCCCACTCTCCCG 379
QY 1985 GCCAAGATGCCAAGATGAGAGGCGCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 2044
Db 378 GCCAAGATGCCAAGATGAGAGGCGCACTCCGACGAGCTTCAAGAACCCATGACCCCTAA 319
QY 2045 ACGGTGCTAATAGTCAGACGCGACAGTGCACCCGAAAGAAACGGCTAGCGCTGATG 2104
Db 318 ACGGTGCTAATAGTCAGACGCGACAGTGCACCCGAAAGAAACGGCTAGCGCTGATG 259
QY 2105 GTGCCAGCTGCAAGCGCGAGCTGCGCTGCACTCAGAAAATCCCTTTGCTAAGGCAACG 2164
Db 258 GTGCCAGCTGCAAGCGCGAGCTGCGCTGCACTCAGAAAATCCCTTTGCTAAGGCAACG 199
QY 2165 GTCTTCCTGGAAG 2178
Db 198 GTCTTCCTGGAAG 185
```

```

RESULT 7
US-09-864-761-14092/c
; Sequence 14092, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14092
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.77
;
US-09-864-761-14092
;
Query Match 8.5%; Score 205; DB 9; Length 489;
Best Local Similarity 100.0%; Pred. No. 1.8e-99;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```



```

QY 312 AGCCCTAGGTGATGGATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTG 371
Db 449 AGCCCTAGGTGATGGATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTG 390
QY 372 TCTTAGTGGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAATTGGGCAATACCTG 431
Db 389 TCTTAGTGGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAATTGGGCAATACCTG 330
QY 432 TTTTGCAATGCAGCACTGAGTGTAACTACACACCACTCTTGGCCAAATTACATGCT 491
Db 329 TTTTGCAATGCAGCACTGAGTGTAACTACACACCACTCTTGGCCAAATTACATGCT 270
QY 492 ATCATGTAACACTCCAAAACATGT 516
Db 269 ATCATGTAACACTCCAAAACATGT 245

```

```

RESULT 8
US-09-864-761-30650/c
; Sequence 30650, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 30650
; LENGTH: 198
; TYPE: DNA

```

```

; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004895.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.77
; OTHER INFORMATION: NT HIT: AB040886.1, EVALUE 2.50e-02
; OTHER INFORMATION: EST_HUMAN HIT: AW504799.1, EVALUE 1.00e-108
; OTHER INFORMATION: SWISSPROT HIT: Q61068, EVALUE 2.00e-13
US-09-864-761-30650

```

```

Query Match      8.2%; Score 198; DB 9; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.1e-95;
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 314 CCTAGGTGATGGATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTGTC 373
Db 198 CCTAGGTGATGGATCGCTCTCCACAGAAAGTTCTTTCCATCTCAGAGATTGTC 139
QY 374 TTAAGTGGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAATTGGGCAATACCTGTT 433
Db 138 TTAAGTGGCAACAACACTCATAGAGTTGGAGCTGGGCTCCAGAATTGGGCAATACCTGTT 79
QY 434 TTGCCAATGCAGCACTGAGTGTAACTACACACCACTCTTGGCCAAATTACATGCTAT 493
Db 78 TTGCCAATGCAGCACTGAGTGTAACTACACACCACTCTTGGCCAAATTACATGCTAT 19
QY 494 CACATGAACACTCCAAAA 511
Db 18 CACATGAACACTCCAAAA 1

```

```

RESULT 9
US-10-027-632-128282/c
; Sequence 128282, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128282
; LENGTH: 504
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-128282

```

```

Query Match      6.5%; Score 157; DB 13; Length 504;
Best Local Similarity 100.0%; Pred. No. 1.6e-73;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	1302	CAGGTC	CCCATGATGTG	AAAAATGG	AGGTGAACTT	TACTCATCC	CACCATAG	CCCCGGCCA	1361
Db	485	CAGGTC	CCCATGATGTG	AAAAATGG	AGGTGAACTT	TACTCATCC	CACCATAG	CCCCGGCCA	426
QY	1362	GTCTCT	CCCGCGCCG	TGCATCAGT	CAGGGTGT	TCCACCA	AAACAGGCT	CGCCGAG	1421
Db	425	GTCTCT	CCCGCGCCG	TGCATCAGT	CAGGGTGT	TCCACCA	AAACAGGCT	CGCCGAG	366
QY	1422	CTTTATCG	GACACAGG	TCCCTCTC	CATGATAA	AG			1458
Db	365	CTTTATCG	GACACAGG	TCCCTCTC	CATGATAA	AG			329

RESULT 10
US-10-027-632-128282/c
/ Sequence 128282, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ Polymorphisms in the Human Genome
/ TITLE OF INVENTION: Polymorphisms in the Human Genome

```

CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 138282

```

	Query Match	6.5%;	Score 157;	DB 16;	Length 504;
	Best Local Similarity	100.0%;	Pred. No. 1.6e-73;		
	Matches 157;	Conservative	0;	Mismatches 0;	Indels 0;
				Gaps	0
QY	1302	CAGGTCCCATGATGTGAAAAATGGAGGTGAACCTTACTCATCCACCCATAGCCCGGCCA	1366		
DB	485	CAGGTCCCATGATGTGAAAAATGGAGGTGAACCTTACTCATCCACCCATAGCCCGGCCA	426		
QY	1362	GTCCCTCTCCCGCCCGGTTCATCAGTCAGCGGGTTCTGCACCAACAACAGGTCGCCCAGG	1421		
DB	425	GTCCCTCTCCCGCCCGGTTCATCAGTCAGCGGGTTCTGCACCAACAACAGGTCGCCCAGG	366		
QY	1422	CTTTATCGACACACAGCTTCCCTCTCACATGATAAG	1458		
DB	365	CTTTATCGACACACAGCTTCCCTCTCACATGATAAG	329		

```

RESULT 11
US-10-027-632-277717
; Sequence 277717, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```

```

, CURRENT APPLICATION NUMBER: US/10/027,632
,
, CURRENT FILING DATE: 2002-04-30
, PRIOR APPLICATION NUMBER: US 60/218,006
,
, PRIOR FILING DATE: 2000-07-12
, PRIOR APPLICATION NUMBER: US 60/198,676
,
, PRIOR FILING DATE: 2000-04-20
, PRIOR APPLICATION NUMBER: US 60/193,483
,
, PRIOR FILING DATE: 2000-03-29
, PRIOR APPLICATION NUMBER: US 60/185,218
,
, PRIOR FILING DATE: 2000-02-24
, PRIOR APPLICATION NUMBER: US 60/167,363
,
, PRIOR FILING DATE: 1999-11-23
, PRIOR APPLICATION NUMBER: US 60/156,358
,
, PRIOR FILING DATE: 1999-09-28
, PRIOR APPLICATION NUMBER: US 60/146,002
,
, PRIOR FILING DATE: 1999-08-09
, NUMBER OF SEQ ID NOS: 325720
, SOFTWARE: FastSeq for Windows Version 4.0
, SEQ ID NO 277717
, LENGTH: 631
, TYPE: DNA
, ORGANISM: Human
,
, US-10-027-632-277717

```

	Query Match	4.0%;	Score 97;	DB 13;	Length 631;
	Best Local Similarity	100.0%;	Pred. NO. 4.4e-41;		
	Matches 97;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2177	AGTTGATGCGCTGTCCTTTGCTGTCCTCCGAGAAGACAAAACTCTTAGAGACCTTCAGGC	2236		
Db	175	AGTTGATGCGCTGTCCTTTGCTGTCCTCCGAGAAGACAAAACTCTTAGAGACCTTCAGGC	234		
QY	2237	TTAGCAACAACTGAAAGGCTCGACGGATGAAATGAG	2273		
Db	235	TTAGCAACAACTGAAAGGCTCGACGGATGAAATGAG	271		

```

RESULT 12
US-10-027-632-277717
; Sequence 277717, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108927.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-01-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-03-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-03-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: fastSEQ for Windows Version 4.0
; SEQ ID NO 277717
; LENGTH: 631
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-277717

```

Query Match 4.0%; Score 97; DB 16; Length 631;
Best Local Similarity 100.0%; Pred. No. 4.4e-41;
Matches 97; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

US-10-027-632-319601/C
; Sequence 319601, Application US/10027632
; Publication NO. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and
; TITLE OF INVENTION: Polymorphisms in
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.6

```

```
Query Match 1.6%; Score 39; DB 16; Length 592;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 TTGTTGTCAGATATTGGAGGATACCTAGATCTAGAG 796
Db 460 TTGTTGTCAGATATTGGAGGATACCTAGATCTAGAG 422

RESULT 18
US-10-258-662-51/c
; Sequence 51, Application US/10258662
; Publication No. US20040082029A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: TANG, Y. Tom
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: BAUGHN, Mariat R.
; APPLICANT: YAO, Monique G.
; APPLICANT: BURFORD, Neil
; APPLICANT: BATRA, Sajeev
; APPLICANT: POLICKY, Jennifer J.
; TITLE OF INVENTION: RNA METABOLISM PROTEINS
; FILE REFERENCE: PF-0771 USN
; CURRENT APPLICATION NUMBER: US/10/258,662
; CURRENT FILING DATE: 2002-10-25
; PRIOR APPLICATION NUMBER: US 60/201,875
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/200,184
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/202,090
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/210,232
; PRIOR FILING DATE: 2000-06-06
; PRIOR APPLICATION NUMBER: US 60/220,553
; PRIOR FILING DATE: 2000-07-25
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PERL Program
; SEQ ID NO 51
; LENGTH: 637
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1647264CB1
US-10-258-662-51

Query Match 1.2%; Score 30; DB 17; Length 637;
Best Local Similarity 100.0%; Pred. No. 7.4e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAATTTAGCTTGA 2412
Db 37 CCCTTTAGTGAGGGTTAATTTAGCTTGA 8

RESULT 19
US-10-149-819-39
; Sequence 39, Application US/10149819
; Publication No. US20030044913A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: TANG, Y. Tom
; APPLICANT: PATTERSON, Chandra
; APPLICANT: BAUGHN, Marian R.
; APPLICANT: LU, Dyung Aina M.
```

```
Query Match 1.6%; Score 39; DB 13; Length 592;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 TTGTTGTCAGATATTGGAGGATACCTAGATCTAGAG 796
Db 460 TTGTTGTCAGATATTGGAGGATACCTAGATCTAGAG 422

RESULT 17
US-10-027-632-319601/c
; Sequence 319601, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319601
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(592)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-319601

Query Match 1.6%; Score 39; DB 13; Length 592;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 758 TTGTTGTCAGATATTGGAGGATACCTAGATCTAGAG 796
Db 460 TTGTTGTCAGATATTGGAGGATACCTAGATCTAGAG 422

RESULT 17
US-10-027-632-319601/c
; Sequence 319601, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 319601
; LENGTH: 592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(592)
; OTHER INFORMATION: n = A,T,C or G
US-10-027-632-319601
```

APPLICANT: SHAH, Purvi
APPLICANT: LAL, Preeti
APPLICANT: AU-YOUNG, Janice
APPLICANT: BURFORD, Neil
TITLE OF INVENTION: EXTRACELLULAR MATRIX AND CELL ADHESION MOLECULES
FILE REFERENCE: PF-0760 PCT
CURRENT APPLICATION NUMBER: US/10/149,819
CURRENT FILING DATE: 2002-06-10
PRIOR APPLICATION NUMBER: 60/172,852; 60/172,354
PRIOR FILING DATE: 1999-12-10; 1999-12-16
NUMBER OF SEQ ID NOS: 42
SOFTWARE: PERL Program
SEQ ID NO 39
LENGTH: 733
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20030044913A1 3683905CB1
US-10-149-819-39

Query Match 1.2%; Score 30; DB 15; Length 733;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 696 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 725

RESULT 20

US-10-313-542-40/c
Sequence 40, Application US/10313542
Publication No. US20030120057A1
GENERAL INFORMATION:
APPLICANT: Roopa, Reddy
APPLICANT: Guegler, Karl, J.
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED P

FILE REFERENCE: PA-0013 US
CURRENT APPLICATION NUMBER: US/10/313,542
CURRENT FILING DATE: 2002-12-05
PRIOR APPLICATION NUMBER: US/09/495,050
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/118,318
PRIOR FILING DATE: 1999-02-01
NUMBER OF SEQ ID NOS: 305
SOFTWARE: PERL Program
SEQ ID NO 40
LENGTH: 847
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20030120057A1 797777CT1
US-10-313-542-40

Query Match 1.2%; Score 30; DB 15; Length 847;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 37 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 8

RESULT 21

US-09-349-015-17
Sequence 17, Application US/09349015A
Patent No. US20020015950A1
GENERAL INFORMATION:
APPLICANT: Jones, Karen Anne
APPLICANT: Volkmut, Wayne

APPLICANT: Walker, Michael
TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
FILE REFERENCE: PB-0013 US
CURRENT APPLICATION NUMBER: US/09/349,015A
CURRENT FILING DATE: 1999-07-07
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 17
LENGTH: 972
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-349-015-17

Query Match 1.2%; Score 30; DB 9; Length 972;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 936 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 965

RESULT 22

US-10-219-664-13
Sequence 13, Application US/10219664
Publication No. US20030129176A1
GENERAL INFORMATION:
APPLICANT: Jones, Karen Anne
APPLICANT: Volkmut, Wayne
APPLICANT: Walker, Michael G.
APPLICANT: Murry, Lynn E.

TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
FILE REFERENCE: PB-0013-1 CIP
CURRENT APPLICATION NUMBER: US/10/219,664
CURRENT FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: 09/349,015
PRIOR FILING DATE: 1999-07-07
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PERL Program
SEQ ID NO 13
LENGTH: 972
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
US-10-219-664-13

Query Match 1.2%; Score 30; DB 15; Length 972;
Best Local Similarity 100.0%; Pred. No. 7.5e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 2412
|||||
Db 936 CCCTTTAGTGAGGGTTAAATTTAGCTTGCA 965

RESULT 23

US-09-880-192-16
Sequence 16, Application US/09880192
Patent No. US20020077470A1
GENERAL INFORMATION:

APPLICANT: Walker, Michael G.
APPLICANT: Volkmut, Wayne
APPLICANT: Klingler, Tod M.
APPLICANT: Azimzai, Valda
TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
FILE REFERENCE: PB-0009-1 CIP
CURRENT APPLICATION NUMBER: US/09/880,192
CURRENT FILING DATE: 2001-06-12
NUMBER OF SEQ ID NOS: 62
SOFTWARE: PERL Program
SEQ ID NO 16
LENGTH: 1112

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 2951269CT1
US-09-880-192-16

Query Match
Best Local Similarity 1.2%; Score 30; DB 9; Length 1112;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db 1075 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1104

RESULT 24
US-10-427-348-16
; Sequence 16, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Valda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 16
; LENGTH: 1112
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 2951269CT1
US-10-427-348-16

Query Match
Best Local Similarity 1.2%; Score 30; DB 15; Length 1112;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db 1075 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1104

RESULT 25
US-09-981-353-187
; Sequence 187, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 187
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 522433CB1

```

```

US-09-981-353-187

Query Match
Best Local Similarity 1.2%; Score 30; DB 9; Length 1273;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db 1238 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1267

RESULT 26
US-09-919-039-191
; Sequence 191, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 191
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 522433CB1
US-09-919-039-191

Query Match
Best Local Similarity 1.2%; Score 30; DB 10; Length 1273;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db 1238 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 1267

RESULT 27
US-10-133-013-59
; Sequence 59, Application US/10133013
; Publication No. US20030166903A1
; GENERAL INFORMATION:
; APPLICANT: Astromoff, Anna
; APPLICANT: Cock, Benjamin G.
; TITLE OF INVENTION: GENES ASSOCIATED WITH VASCULAR DISEASE
; FILE REFERENCE: PA-0049 US
; CURRENT APPLICATION NUMBER: US/10/133,013
; CURRENT FILING DATE: 2003-04-25
; PRIOR APPLICATION NUMBER: 60/287,067
; PRIOR FILING DATE: 2001-04-27
; NUMBER OF SEQ ID NOS: 271
; SOFTWARE: PERL Program
; SEQ ID NO 59
; LENGTH: 1273
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030166903A1 522433CB1
US-10-133-013-59

Query Match
Best Local Similarity 1.2%; Score 30; DB 15; Length 1273;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412

```

```
Db 1238 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 1267
|||||
RESULT 28
US-09-349-015-2
; Sequence 2, Application US/09349015A
; Patent No. US20020015950A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013 US
; CURRENT APPLICATION NUMBER: US/09/349,015A
; CURRENT FILING DATE: 1999-07-07
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1702
; TYPE: DNA
; ORGANISM: HOMO SAPIENS
US-09-349-015-2
Query Match 1.2%; Score 30; DB 9; Length 1702;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 1665 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 1694
|||||
RESULT 29
US-10-219-664-2
; Sequence 2, Application US/10219664
; Publication No. US20030129176A1
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: ATHEROSCLEROSIS-ASSOCIATED GENES
; FILE REFERENCE: PB-0013-1 CIP
; CURRENT APPLICATION NUMBER: US/10/219,664
; CURRENT FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: 09/349,015
; PRIOR FILING DATE: 1999-07-07
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 1702
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
US-10-219-664-2
Query Match 1.2%; Score 30; DB 15; Length 1702;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 1665 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 1694
|||||
RESULT 30
US-09-880-192-32/c
; Sequence 32, Application US/09880192
; Patent No. US20020077470A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Jones, Karen Anne
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-1 CIP
; CURRENT APPLICATION NUMBER: US/09/880,192
; CURRENT FILING DATE: 2001-06-12
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020077470A1 3795510CT1
US-09-880-192-32
Query Match 1.2%; Score 30; DB 9; Length 1833;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 37 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 8
|||||
RESULT 31
US-10-427-348-32/c
; Sequence 32, Application US/10427348
; Publication No. US20030175795A1
; GENERAL INFORMATION:
; APPLICANT: Walker, Michael G.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Klingler, Tod M.
; APPLICANT: Azimzai, Yalda
; TITLE OF INVENTION: POLYNUCLEOTIDES ASSOCIATED WITH CARDIAC MUSCLE FUNCTION
; FILE REFERENCE: PB-0009-2 CON
; CURRENT APPLICATION NUMBER: US/10/427,348
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 09/880,192
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: US 09/299,708
; PRIOR FILING DATE: 1999-04-26
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030175795A1 3795510CT1
US-10-427-348-32
Query Match 1.2%; Score 30; DB 15; Length 1833;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2383 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 2412
Db 37 CCCTTAGTGAGGGTTAAATTTAGCTTGCA 8
|||||
RESULT 32
US-09-971-392-216/c
; Sequence 216, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Pearson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
```

```
FILE REFERENCE: PA-0029 US
CURRENT APPLICATION NUMBER: US/09/971,392
CURRENT FILING DATE: 2001-10-03
PRIOR APPLICATION NUMBER: 60/237,652
PRIOR FILING DATE: 2000-10-03
NUMBER OF SEQ ID NOS: 260
SOFTWARE: PERL Program
SEQ ID NO 216
LENGTH: 1926
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Template ID: 1095059.19
US-09-971-392-216

Query Match      1.2%; Score 30; DB 10; Length 1926;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db      36 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 7

RESULT 33
US-09-818-143-20/c
Sequence 20, Application US/09818143
Patent No. US20020019000A1
GENERAL INFORMATION:
APPLICANT: Walker, Michael G.
APPLICANT: Volkmut, Wayne
APPLICANT: Klingler, Tod M.
TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES
FILE REFERENCE: PB-0004 CIP
CURRENT APPLICATION NUMBER: US/09/818,143
CURRENT FILING DATE: 2001-03-26
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PERL Program
SEQ ID NO 20
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 3948614CB1
US-09-818-143-20

Query Match      1.2%; Score 30; DB 9; Length 1987;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db      39 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 10

Search completed: August 17, 2004, 12:08:22
Job time : 3508 secs
```

```
FILE REFERENCE: PA-0029 US
CURRENT APPLICATION NUMBER: US/09/971,392
CURRENT FILING DATE: 2001-10-03
PRIOR APPLICATION NUMBER: 60/237,652
PRIOR FILING DATE: 2000-10-03
NUMBER OF SEQ ID NOS: 260
SOFTWARE: PERL Program
SEQ ID NO 216
LENGTH: 1926
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Template ID: 1095059.19
US-09-971-392-216

Query Match      1.2%; Score 30; DB 10; Length 1926;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db      36 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 7

RESULT 33
US-09-818-143-20/c
Sequence 20, Application US/09818143
Patent No. US20020019000A1
GENERAL INFORMATION:
APPLICANT: Walker, Michael G.
APPLICANT: Volkmut, Wayne
APPLICANT: Klingler, Tod M.
TITLE OF INVENTION: POLYNUCLEOTIDES COEXPRESSED WITH MATRIX-REMODELING GENES
FILE REFERENCE: PB-0004 CIP
CURRENT APPLICATION NUMBER: US/09/818,143
CURRENT FILING DATE: 2001-03-26
NUMBER OF SEQ ID NOS: 23
SOFTWARE: PERL Program
SEQ ID NO 20
LENGTH: 1987
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 3948614CB1
US-09-818-143-20

Query Match      1.2%; Score 30; DB 9; Length 1987;
Best Local Similarity 100.0%; Pred. No. 7.7e-05;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2383 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 2412
Db      39 CCCTTTAGTGAGGGTTAAATTTTAGCTTGCA 10

Search completed: August 17, 2004, 12:08:22
Job time : 3508 secs
```